

FOR OFFICIAL USE ONLY

ACCESS DB #

42000

PLEASE PRINT CLEARLY

Location (Bldg/Room#): CM1-7E15

Mail Box CM1-7E12

Scientific and Technical Information Center

SEARCH REQUEST FORM

m:

Date: 8 May 01 Requester's Full Name: _____ Examiner #: S. DEVI

Art Unit: 1645 Phone (308) 9347 Serial Number: 09/151,409

Results Format Preferred (circle): PAPER DISK E-MAIL

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): JAMES B. DALE

Earliest Priority Date: 9.12.97

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the cited species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

For Sequence Searches Only* Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with appropriate serial number.

Please ask MS. BEVERLY SHEARS to perform this search.

Please see attached claims with key words highlighted and/or Examples and synonyms provided.

Please include the following databases: Embase, Medline, Biosis, CA (Dialog 50), JAPIO, JICTEplus, Dialog 35, 65, 77, 144, 256, 266, 440, 348, 357, 113, 129, 130, 156 and 60.

Please perform an inventor's name search.

Point of Contact:

Beverly Shears

Technical Info. Specialist

CM1 12C14 Tel: 308-4994

NOT a
sequence
search

Please perform a sequence and
an interference search for
SEQ ID NO. 15 and 16.

Thank you. ☺

Please return the attached claims and this sequence
search request form along with the search reports.



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:42:35 ; Search time 1714.27 seconds
(without alignments)
9319.598 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagcatggtgcgcagtag,.....aggcagcgaagtaaaagctt 1158
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13168883 seqs, 6898233199 residues
Total number of hits satisfying chosen parameters: 26337766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA_Main:*

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US098_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US099_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1158	100.0	1158	15	US-09-151-409-15	Sequence 15, Appl
2	419	36.2	741	3	US-07-945-954A-2	Sequence 2, Appl
3	328.2	28.3	822	3	US-07-945-954A-3	Sequence 3, Appl
4	301.8	26.1	741	3	US-07-945-954A-6	Sequence 6, Appl
5	301.8	26.1	1029	3	US-07-945-954A-8	Sequence 8, Appl
6	241	20.8	861	3	US-07-945-954A-1	Sequence 1, Appl
7	239.4	20.7	522	3	US-07-945-954A-7	Sequence 7, Appl
8	151	13.0	1332	7	US-08-325-278-5	Sequence 5, Appl
9	124.4	10.7	561	3	US-07-945-954A-4	Sequence 4, Appl
10	82.6	7.1	7218	8	US-08-466-194-14	Sequence 10, Appl
11	76.2	6.6	408	3	US-07-945-954A-10	Sequence 13166, A
12	73.2	6.3	1446	18	US-09-404-520-13166	Sequence 2159, Ap
13	72.4	6.3	957	18	US-09-457-659-2159	Sequence 5964, Ap
14	72.4	6.3	937	25	US-09-652-127-5964	Sequence 2729, Ap
15	70.4	6.1	989	17	US-09-397-761A-2729	Sequence 2719, Ap
16	70	6.0	889	17	US-09-397-761A-2719	Sequence 411, App
17	70	6.0	889	25	US-09-641-529-411	Sequence 3082, Ap
18	70	6.0	924	17	US-09-397-761A-3082	Sequence 362, App
19	68	5.9	2943	1	PCT-US99-26796-362	Sequence 2721, Ap
20	67.4	5.8	931	17	US-09-397-761A-2721	Sequence 674, App
21	67.4	5.8	931	25	US-09-641-529-674	Sequence 2728, Ap
22	67.2	5.8	985	17	US-09-397-761A-2728	Sequence 2722, Ap
23	66.4	5.7	857	17	US-09-397-761A-2722	Sequence 444, App
24	66.4	5.7	857	25	US-09-641-529-444	Sequence 13069, A
25	66.2	5.7	948	18	US-09-404-520-13069	Sequence 20050, A
26	66.2	5.7	994	52	US-60-202-214-20050	Sequence 13290, A
27	66.2	5.7	1449	18	US-09-404-520-13290	Sequence 2735, Ap
28	65.4	5.6	979	17	US-09-397-761A-2735	Sequence 2715, Ap
29	65.2	5.6	972	17	US-09-397-761A-2715	Sequence 680, App
30	65.2	5.6	972	25	US-09-641-529-680	Sequence 1953, Ap
31	65.2	5.6	1131	29	US-09-739-449-1953	Sequence 16094, A
32	65	5.6	983	28	US-09-705-926-16094	Sequence 376, App
33	64.4	5.6	786	17	US-09-397-761A-2720	Sequence 232, App
34	64.4	5.6	786	25	US-09-641-529-376	Sequence 13295, A
35	64.2	5.5	3579	1	PCT-US99-26796-232	Sequence 8012, Ap
36	63.8	5.5	1167	18	US-09-404-520-13295	Sequence 1335, Ap
37	63.2	5.5	723	29	US-09-739-449-8012	Sequence 3635, Ap
38	63.2	5.5	892	29	US-09-739-449-1335	Sequence 345, App
39	63.2	5.5	905	25	US-09-652-816-3635	Sequence 354, App
40	63.2	5.5	1998	1	PCT-US99-26796-345	Sequence 13305, A
41	62.8	5.4	834	52	US-60-207-458-354	Sequence 13441, A
42	62.6	5.4	1193	18	US-09-404-520-13305	Sequence 2158, Ap
43	62	5.4	877	51	US-60-196-868-13441	Sequence 5963, Ap
44	61.4	5.3	783	18	US-09-457-659-2158	
45	61.4	5.3	783	25	US-09-652-127-5963	

Query Match	28.3%	Score 328.2;	DB 3;	Length 822;
Best Local Similarity	87.6%;	Pred. No. 1.4e-59;		
Matches 404;	Conservative	0;	Mismatches 3;	Indels 54;
				Gaps 2;


```
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1029
;
US-07-945-954A-8

Query Match      26.1%; Score 301.8; DB 3; Length 1029;
Best Local Similarity 68.3%; Pred. No. 5.8e-54;
Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;

QY 486 gagagtcggttatactagcaccaggaagataagctaaaaaattattgacgatct 545
Db 3 GAGAGTCGTTTACTAGGATACCCAGAGATAAGCTAAAAAAATATTGACGATCT 62
QY 546 tgacgcaaaagaacatgaattacaacaacagaatgagagttatctctgcagaacgtga 605
Db 63 TGACGCAAAAGAACAATGAATTACACAACACAGAATGAGAGTTATCTGGATCCAGAGTGT 122
QY 606 tggtaactctaggaagtattagaagatcttcgacgaacaatcccgcaatcacaaaatat 665
Db 123 TCCTAGGGGGACGGTAGAAAAACCCCG-----ACAAAGC 155
QY 666 acgtttactgcacgaaacaggacttaaaagcgagattagagatgcaatggaagtgc 725
Db 156 ACGAGAAGCTTCTTAACAAGTATGACGTAGACAACTCTATGTTACAAGCTAATAATGACAA 215
QY 726 aggaagagatttttaagagagctggttaccttctgttagatcaggttacacaattatatactaa 785
Db 216 CTATGTCGACGCCGTGACTAGGGGTACAATTAATGACCCGCAAGAGCAAAAGAGCTCT 275
QY 786 acataatagtaattaccacaataataatgcacaagctggcagacttgacctgagacaaa 845
Db 276 TCACAAGTATGAGCTAGAAAACCATGACTTTAAAAAATAAGAATGAAGGGTTAAAAAAGTGA 335
QY 846 ggcgtgaatatctaaagcccttaagtgattggctgagagctgttacagagttaaatat 905
Db 336 GAATGAAGGGTTAAAAAAGTGAATGAAGGGTTAAAAAAGTGAAGGGTTAAAAAAGTGAAGGGTT 395
QY 906 cga-----tgctcgactaggtctcagacagatactcttgaaaaagttacaagaactgc 959
Db 396 TGACCCATGGTCCGACTAGGCTTCAGACAGATACCTCTGGAAGAAAGTACAAGAACTGC 455
QY 960 tgacaagtttgagatagaaacaatacgtttaaaacttaaaacttaaaacttaaaacttaaaacttaaa 1019
Db 456 TGACAAGTTTGAGATAGAAAACAAATAGCTTTAAAACTTTAAGAACTAGTGAAGTTTAA 515
QY 1020 taataaagcgttaaaagatcataatgatgagtttaactgaagagttgagtaagtctaaga 1079
Db 516 TAATAAAGCGTTAAAGATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 575
QY 1080 gaaactcgtaaaaatgataaatacctatctgaaaaagctagtaaaattcaagaattaga 1139
Db 576 GAACTACGTAAAAATGATAAATCACTATCTGAAAAAGCTAGTAAAAATCAAGAATTAGA 635
QY 1140 ggcacgttaagtaaaagctt 1158
Db 636 GGCACGTAAGGCTGATCTT 654
```

```
RESULT 6
US-07-945-954A-1
; Sequence 1, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..861
;
US-07-945-954A-1
```

```
Query Match      20.8%; Score 241; DB 3; Length 861;
Best Local Similarity 98.0%; Pred. No. 4.2e-41;
Matches 244; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 910 gtcgcgactaggtctcagacagatactctggaagaagtacagaacgtgctgacaagttt 969
Db 4 GTCGCGACTAGGTCTCAGACAGATACCTCTGGAAGAAAGTACAAGAACGTCGACAAAGTTT 63
QY 970 gagatagaaaaacaatacgtttaaaacttaagaatagtagtacttaagttttaataaataaagcg 1029
Db 64 GAGATAGAAAAACAATACGTTAAACCTTAAGAAATAGTACTTAAGTTTAAATAATAAAGCG 123
QY 1030 ttaagaatcataatgatgatgagttgaactgaagagttgagtaagtctaaagagaaactacgt 1089
Db 124 TTAAGAAGATCATAATGATGAGTTAACTGAAGAGTTGAGTAACTGAAGAGAAACTACTAGT 183
QY 1090 aaaaatgataaatacctatctgaaaaaagctagtaaaattcaagaattagagggcacgtaag 1149
Db 184 AAAAATGATAATCACTATCTCAAAAAGCTAGTAAATTCAGAAATTCAGAGCAGGTAAAG 243
QY 1150 taaaagctt 1158
Db 244 GCTGATCTT 252
```

Db 244 GCTGATCTT 252

RESULT 8

US-08-325-278-5

; Sequence 5, Application US/08325278

; GENERAL INFORMATION:

; APPLICANT: Bjvrck, Lars

; APPLICANT: Sjvbring, Ulf

; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/325,278

; FILING DATE: 26-OCT-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: McMasters, David D.

; REGISTRATION NUMBER: 33,963

; REFERENCE/DOCKET NUMBER: 450023.401

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1332 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1329

; US-08-325-278-5

Query Match 13.0%; Score 151; DB 7; Length 1332;

Best Local Similarity 100.0%; Pred. No. 5.3e-22;

Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 598 aacggtgatggttaactcttaggaagattatagaagatcttgcagcaaaacaccccgcaata 657

|||||

Db 1 AACGGTGATGGTAACTCTTAGGGAAGTTATAGAAGATCTTGACGAAACAAATCCCGCAATA 60

QY 658 caaatatagctttacgtcacgaaacaaagagacttaaaagcgagattagagaatgcaatg 717

|||||

Db 61 CAAATATATACGTTTACGTCACCAACAAAGGACTTAAAGCGAGATTAGAGATGCAATG 120

QY 718 gaagttgcaggagagattttaagagagctg 748

|||||

Db 121 GAAGTTGCAGGAAGAGATTTTAAGAGAGCTG 151

RESULT 9

US-07-945-954A-4

; Sequence 4, Application US/07945954A

; GENERAL INFORMATION:

; APPLICANT: Dale, James B.

; APPLICANT: Lederer, James W.

; TITLE OF INVENTION: Recombinant Multivalent M Protein

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weiser & Associates

; STREET: 230 South Fifteenth Street, Suite 500

; CITY: Philadelphia

; STATE: Pennsylvania

; COUNTRY: U.S.A.

; ZIP: 19102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/945,954A

; FILING DATE: 16-SEP-1992

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Weiser, Gerard J.

; REGISTRATION NUMBER: 19,763

; REFERENCE/DOCKET NUMBER: 372.5672P

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-875-8383

; TELEFAX: 215-875-8394

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 522 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Streptococcus pyogenes

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..519

; US-07-945-954A-7

Query Match 20.7%; Score 239.4; DB 3; Length 522;

Best Local Similarity 97.6%; Pred. No. 8.3e-41;

Matches 243; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 910 gtccgactaggtctcagacagatctctggaagaaagacaaagcgtgctgcaagttt 969

|||||

Db 4 GTCGCGACTAGGTCTCAGACAGATACCTCTGGAAGAAAGTACAAAGCGTCTGACAAGTTT 63

QY 970 gagatagaaacaatacgttaaaacttaagaatagtgacttaagttttaataataaagcg 1029

|||||

Db 64 GAGATAGAAAACAATACGTTTAAACTTAAGAAATAGTGACTTAAGTTTATATAAAGCG 123

QY 1030 ttaaagatcataatgatgagttaactgaaggttgagtaatgctaaagagaaactacgt 1089

|||||

Db 124 TTAAGATCATAATGATGAGTTAACTTAACGAAGTTGAGTAATGCTAAAGAGAAACTACGT 183

QY 1090 aaaaatgataactactctcgaagaaagcgtagtaaaattcaagaattagagcgacgttaag 1149

|||||

Db 184 AAAAAATGATAATCACTACTCTCAAAAAGCTAGTAAAAATCAAGAATTAGAGGCACGTAAG 243

QY 1150 taaaagcgtt 1158

|||||

Qy 175 aaagagaactaogtgatccgcgtgactagggtacataaatgaccgccaaagaca 234
| | | | | : : : : : : : : : : : : : : : :
Db 1451 ATAGACGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 139

Qy 367 gaggtcac-----agagdtttctagggggcggtagaaacccggacaaacacga 420
|| |
Idb 361 CCGGACAAACGACGAGAGTGTTTCTTAGGGGGACGTAGAAACCCGGACAAACGACGA 420

RESULT 11
 S-07-945-954A-10
 Sequence 10, Application US/07945954A
 GENERAL INFORMATION:
 APPLICANT: Dale, James B.
 APPLICANT: Lederer, James W.
 TITLE OF INVENTION: Recombinant Multivalent M Protein
 TITLE OF INVENTION: Vaccine
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Weiser & Associates
 STREET: 230 South Fifteenth Street, Suite 500
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: U.S.A.
 ZIP: 19102
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/945,954A
 FILING DATE: 16-SEP-1992
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: Weiser, Gerard J.
 REGISTRATION NUMBER: 19,763
 REFERENCE/DOCKET NUMBER: 372.5672P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-875-8383
 TELEFAX: 215-875-8394
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 408 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Streptococcus pyogenes
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..405
 S-07-945-954A-10

Query Match	6.3%	Score 73.2	DB 18	Length 1446
Best Local Similarity	44.0%	Pred. No. 1.6e-05		
Matches 493	Conservative	0	Mismatches 618	Indels 9
Gaps				
QY	41	aaaaagtacaagaacgctgcgacaagtttgagatagaacaacatacgtttaaacactaaga	100	
DB	285	aaaaataaaaaatacaaaagataagaaaaataaagaataataaattcaaaaataaaa	344	
QY	101	a--tagtgacttaagttttataataaaaagcgttaaaagatcataatgatgatgaactga	158	
DB	345	aattataggaanaaaaaataaaaaattataataaaaaaanaaaaaaagaataataa	404	
QY	159	agagttgagtaatgctaaagagaaactcactgctcgccgctgactcaggggtacataaaa	218	
DB	405	aaaaataataaagagggaanaaaaaataaagagaaaaataaagaanaataaaaaataataa	464	
QY	219	tgaccgcgaagagcaaaagagctcttgacaagtatgagctagaaaaccatgacttaaa	278	
DB	465	taaaaagaaaaaaatgaataataaagaattgaaaaataataagaataagaaaata	524	
QY	279	aactaagaatatgaaggggttaaaaactggaatatgaaggttataaacicgagaatgaaggggt	338	
DB	525	atataagaaaaaaaggttaaaaaatatagaaataaaaaataaataagaaaaataagaaaaata	584	
QY	339	aaaaactcgagaatgaaggggttaaaaactcgaggtcgacagagtggtttcttctaagggggacggt	398	
DB	585	aaattataaaaataaattatagataataaagaanaaagaatgatatataaaaataataa	644	
QY	399	agaaacccggacaaagacgcagagaactctcttaacaagtatgacgtgagagacctctatgtt	458	
DB	645	ataaaacagagaaaaaagaaaaataaaaaaagaatataataatacaacaataagaatata	704	

Search completed: May 13, 2001, 05:29:32
Job time: 2817 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:42:05 ; Search time 88.15 Seconds
(without alignments)
4928.574 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcacgtggtgcgactag.....aggcagcgaagtaaaagctt 1158

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262727 seqs, 187588001 residues

Total number of hits satisfying chosen parameters: 525454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents.NA.New.*

- 1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
- 2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
- 3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
- 4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
- 5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
- 6: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.2	5.6	1131	5	US-09-739-449-1953
2	63.2	5.5	723	5	US-09-739-449-8012
3	63.2	5.5	892	5	US-09-739-449-1335
c 4	59.6	5.1	1250	5	US-09-739-449-1859
5	58.2	5.0	43546	6	US-60-254-168-38
c 6	56.6	4.9	798	5	US-09-739-449-2641
7	56.4	4.9	909	5	US-09-739-449-636
c 8	54.6	4.7	543	5	US-09-739-449-8029
c 9	53.4	4.6	834	5	US-09-739-449-2256
10	52.4	4.5	902	5	US-09-739-449-1929
11	52	4.5	1167	5	US-09-739-449-2034
c 12	50.8	4.4	1040	5	US-09-739-449-2287
c 13	50.8	4.4	147068	6	US-60-248-505-357
c 14	50.8	4.4	334854	6	US-60-248-505-28
c 15	50.6	4.4	937	5	US-09-739-449-2534
16	50.4	4.4	845	5	US-09-801-833-4847
c 17	50	4.4	1006	5	US-09-739-449-621
18	48.8	4.2	780	5	US-09-739-449-778
c 19	48.4	4.2	757	5	US-09-739-449-2286
c 20	47.8	4.1	753	5	US-09-739-449-1234
c 21	47	4.1	944	5	US-09-739-449-492
c 22	46.8	4.0	107487	6	US-60-248-505-231
c 23	46	4.0	787	5	US-09-739-449-2501
c 24	46	4.0	1418	5	US-09-739-449-1613
c 25	44.8	3.9	684	5	US-09-801-833-5354
26	44.2	3.8	499	5	US-09-739-449-1332
27	44	3.8	2795	1	PCT-US01-04098A-2166

28	44	3.8	3058	1	PCT-US01-04098A-198	Sequence 198, App
c 29	43.8	3.8	776	5	US-09-739-449-2257	Sequence 2257, App
c 30	43.8	3.8	4017	5	US-09-783-514-1895	Sequence 1895, App
31	43.6	3.8	4704	5	US-09-801-833-7699	Sequence 7699, App
c 32	43	3.7	681	5	US-09-739-449-2434	Sequence 2434, App
c 33	42.8	3.7	778	5	US-09-739-449-2435	Sequence 2435, App
c 34	42.6	3.7	345	5	US-09-739-449-8021	Sequence 8021, App
c 35	42.4	3.7	3491	4	US-08-956-171C-117	Sequence 117, App
36	42.2	3.6	987	5	US-09-739-449-719	Sequence 719, App
c 37	42	3.6	629	5	US-09-739-449-1185	Sequence 1185, App
c 38	42	3.6	988	5	US-09-739-449-2009	Sequence 2009, App
c 39	42	3.6	16596	1	PCT-US01-01351-504	Sequence 504, App
40	41.8	3.6	719	5	US-09-739-449-2589	Sequence 2589, App
41	41.8	3.6	1181	5	US-09-739-449-273	Sequence 273, App
c 42	41.8	3.6	5768	1	PCT-US01-01339-8019	Sequence 8019, App
c 43	41.8	3.6	5768	1	PCT-US01-01329-2852	Sequence 2852, App
44	41.6	3.6	315	5	US-09-739-449-366	Sequence 366, App
45	41.6	3.6	548	5	US-09-739-449-1457	Sequence 1457, App

ALIGNMENTS

RESULT 1
US-09-739-449-1953
; Sequence 1953, Application US/09739449
; GENERAL INFORMATION: Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 1953
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1131)
; OTHER INFORMATION: unsure at all n locations
US-09-739-449-1953

Query Match 5.6%; Score 65.2; DB 5; Length 1131;
Best Local Similarity 41.6%; Pred. No. 2.4e-05;
Matches 342; Conservative 0; Mismatches 479; Indels 1; Gaps 1;

QY	41	aaaaagtacaaacgtgtgcacaaagtttgagatagataaaacatacgttaaaacttaaga	100
Db	144	aaaaaaaaaaganaaaaaaagaaaaaagaaaaaannnnnnnaaaaaa	203
QY	101	atagtacttaagttttaataataaagcgttaaaagtcataatgatgactgaag	160
Db	204	aaaaanaaaaaaaggggaaaaaagaaaaaagaaaaaannnnnaaaangaa	263
QY	161	agttgtaagttaagagaaactactggtggtcccgtagctgggtacataataatg	220
Db	264	aaaaaaaaaagaaaaaagaaaaaagaaaaaaggggaaaaaagaaaaa	323
QY	221	accgcacaaagacaaagagctcttgacaaagtatagctagaaacccatgacttaaaa	280
Db	324	annnaaaaaaagaaaaaagaaaaaagaaaaaannnnnaaaaaaagaaaaa	383
QY	281	ctaagaatgaaggtttaaaactgagaatgaaggggttaaaactgagaatgaaggttaa	340
Db	384	aaaaaagaaaaaagaaaaaagaaaaaagaaaaaannnnnaaaaaaagaaaaa	443
QY	341	aaactgagaatgaaggggttaaaactgagtcgacagtggtttcttaggggacggttag	400

Tue May 15 07:27:15 2001

```
QY 581 aqaagtctctgcagaacggtgatggttaactctcctaggggaagtattagaagctctgcag 640
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 266 AAAAAAAAAAAAAAAAAAAGCCGCCGCCCAAAAAAAAAAAAAAAAAAATAAATAA 207
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 641 caaacaatccgcgaatacaataatcgtttacgtcacgaaacaaggacttaaaagc 698
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 206 AAAAAAAAAAAAAAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAAC 149

RESULT 10
US-09-739-449-1929
; Sequence 1929, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 1929
; LENGTH: 902
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(902)
; OTHER INFORMATION: unsure at all n locations
US-09-739-449-1929
```

```
Query Match 4.5%; Score 52.4; DB 5; Length 902;
Best Local Similarity 32.7%; Pred. No. 0.011;
Matches 209; Conservative 0; Mismatches 430; Indels 0; Gaps 0;

QY 516 agaatgctaaataattatgcgtctctgcgcaaaagacatgaattacaaca 575
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 aaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 576 gaatgagattctctgcagacggtgatggttaactcctagggaggttatgagatct 635
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 636 tgcagcaaatccgcgaatacaataatactgttctgcacgaaacaaggacttaa 695
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 ggggggnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 696 agcgagattagaaatgcaatggaagtgcggaagagattttaagagagctggtacctt 755
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 aaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 756 gtagatcaggtttacacattatactaacataatagtaattaccacaataatgc 815
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 annnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 462
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 816 acaagctggcagactgcctgcagacaaaggctgaatatctaaaggcccttaatgattg 875
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 463 aaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 522
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 876 ggctgagagctgttacaagagtttaatatctgctgcgactaggtctcagacagatc 935
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 523 aaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 582
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 936 tctggaaaaatcacgaacgtgctgacaaagtttgatagatagaaacaatactgttaaaact 995
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 583 anaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 642
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 996 taagaatagtgacttaagttttaataataaagcgttaaaagatcataatgatgagttaac 1055
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 643 aaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 702
```

```
QY 1056 tgaagagttgagtaagtctaaagagaactacgttaaaaaatgataaatactactctgaaaa 1115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 703 annaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 762
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1116 agctagtaaaattcaagaattagtaggcacgttaagtaaaa 1154
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 aaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 801

RESULT 11
US-09-739-449-2034
; Sequence 2034, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 2034
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1167)
; OTHER INFORMATION: unsure at all n locations
US-09-739-449-2034
```

```
Query Match 4.5%; Score 52; DB 5; Length 1167;
Best Local Similarity 42.4%; Pred. No. 0.014;
Matches 172; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY 41 aaaaagacaagaacgtgctgacaaagtttgatagatagaaaaacaatactgttaaaacttaaga 100
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 674 aaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 733
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 101 atagtgcacttaagttttaataataaagcgtttaaagatcataatgatgagtttaactgaag 160
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 734 aaaaagaagaagagaaaaagaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnn 793
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 161 agttgagatgctaaagagaactacgtggtacgcgcgtgactgaggggtacataataatg 220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 794 anaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 853
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 221 accgcgaagagcaaaagaaagctcttcgacaagatgatgtagaaaccatgacttaaaaa 280
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 854 aaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 913
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 281 ctaagaatgaaggggttaaaaaactgagaatgaaggggttaaaaaactgagaatgaaggggttaa 340
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 914 aagagaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 973
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 341 aaactgagaatgaaggggttaaaaaactgaggtcgacagagagtttctcctgaggggacggtag 400
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 974 naaaaaaaagaaaaagagaagagaggnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1033
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 401 aaaccgcgacaaagcgcgagactctttaacagatgacgtaga 446
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1034 aaaaagnnnngaagaaaaaaannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn 1079
```

```
RESULT 12
US-09-739-449-2287/c
; Sequence 2287, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
```


[illegible]

Search completed: May 13, 2001, 05:03:55
Job time: 1310 sec

	Query Match	4.4%	Score 50.6;	DB 5;	Length 937;
	Best Local Similarity	44.1%;	Pred. No. 0.027,		
	Matches 212;	Conservative	0; Mismatches 269;	Indels	Gaps
Qy	654	aatacaaaataatcgttttacgctcacgaacaacagactttaaagcgagattagagaatgc	713		
Dd	935	AAGAAAAAAGAAAAAAAAAAAAAAAAAATTTAAAGAAAAGAAATGAAATAAAAAAAAAAGTAA	876		
Qy	714	aatggaagtgcaggaaagattttaagagagcigtgcccttgtagtcaggttacaca	773		
Dd	875	AATTAAAAAANAANNTAATAATACAAAAAAMGAAGAANAATGTAAAAAANAATTAAGAAG	816		
Qy	774	attatatctaacaataatagtaattaccaacaataataatgcacaagctggcgagacttga	833		
Dd	815	AATAAAAAAAAAAATAAAAAATAATATAAATAATATAAAAAAGAAATAAACATCAAAAAAATAAA	756		
Qy	834	cctgagacaaagccgtgaatcatctaaagacccttaatgatitggctgagaggcttgttaca	893		
Dd	755	AAAAAGAAAAAANAATATAAAGAAAAACAACAAAGATATAAAAAAAGAAAAAGAAAAA	696		
Qy	894	agadgttaaatatcgatgctgcgactaggtcttcagacacatactctggaaaaagtcacaaga	953		
Dd	695	AAAAATAGAAAAAANAANTGTTTTAAAAAANAATTAGATATAATTAAGAAAAAANAATTAATAAT	636		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 05:48:01 ; Search time 1719.94 Seconds
(without alignments)
9288.875 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcatcgatggtcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13168883 seqs, 6898233199 residues
Word size: 10
Total number of hits satisfying chosen parameters: 6066217

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending_Patents_NA_Main.*
1: /cgn2_6/ptodata/2/pna/pna/US06_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/pna/US07_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/pna/US08_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/pna/US09_COMB.seq.*

Result No.	Score	Query Match %	Length	ID	Description
1	1158	100.0	1158	15	US-09-151-409-15
2	402	34.7	741	3	US-07-945-954A-2
3	240	20.7	861	3	US-07-945-954A-1
4	217	18.7	522	3	US-07-945-954A-7
5	217	18.7	741	3	US-07-945-954A-6
6	217	18.7	822	3	US-07-945-954A-3
7	217	18.7	1029	3	US-07-945-954A-8
8	151	13.0	1332	7	US-08-325-278-5
9	53	4.6	561	3	US-07-945-954A-4
10	52	4.5	204	3	US-07-945-954A-9
11	52	4.5	408	3	US-07-945-954A-10
12	52	4.5	918	3	US-07-945-954A-5
13	48	4.1	765	3	US-07-945-860A-2
14	48	4.1	765	3	US-07-945-860B-2
15	48	4.1	855	13	US-08-914-479-3
16	48	4.1	855	3	US-07-945-860A-3
17	48	4.1	855	3	US-07-945-860B-3
18	48	4.1	855	13	US-08-914-479-5
19	36	3.1	417	3	US-07-945-860A-1
20	36	3.1	417	3	US-07-945-860B-1
21	36	3.1	417	13	US-08-914-479-1
22	27	2.3	33	15	US-09-151-409-14
23	25	2.2	30	15	US-09-151-409-5
24	25	2.2	36	15	US-09-151-409-1
25	24	2.1	30	15	US-09-151-409-2
26	24	2.1	30	15	US-09-151-409-3
27	24	2.1	30	15	US-09-151-409-4
28	24	2.1	30	15	US-09-151-409-6
29	24	2.1	30	15	US-09-151-409-8
30	24	2.1	30	15	US-09-151-409-9
31	24	2.1	30	15	US-09-151-409-10
32	24	2.1	30	15	US-09-151-409-11
33	24	2.1	30	15	US-09-151-409-12
34	24	2.1	30	15	US-09-151-409-13
35	24	2.0	37066	56	US-60-243-468-190
36	22	1.9	32	14	US-09-095-028-3
37	22	1.8	351	22	US-09-572-409-39350
38	21	1.8	30570	28	US-09-702-134-29869
39	21	1.8	30763	24	US-09-620-392-11277
40	21	1.7	286	24	US-09-837-086A-35904
41	20	1.7	405	19	US-09-521-640-265461
42	20	1.7	405	46	US-60-140-769-50892
43	20	1.7	612	25	US-09-654-617-9386
44	20	1.7	612	27	US-09-684-016-9386
45	20	1.7	1524	25	US-09-654-617-194272
46	20	1.7	1524	27	US-09-684-016-194272
47	20	1.7	1524	27	US-09-684-016-194272

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

48	20	1.7	8611	18	US-09-404-520-5819	Sequence 5819, Ap	C 121	19	1.6	4417	27	US-09-698-012-8398	Sequence 8398, Ap
49	20	1.7	32768	52	US-60-200-382-31	Sequence 31, Appl	C 122	19	1.6	4417	27	US-09-698-014-5331	Sequence 5331, Ap
50	20	1.7	32768	52	US-60-200-382-32	Sequence 32, Appl	C 123	19	1.6	4417	27	US-09-699-997-11577	Sequence 11577, Ap
51	20	1.7	32768	52	US-60-200-382-33	Sequence 33, Appl	C 124	19	1.6	4417	27	US-09-699-998-9374	Sequence 9374, Ap
52	20	1.7	32768	53	US-60-212-659-191	Sequence 191, App	C 125	19	1.6	4417	28	US-09-710-285-2019	Sequence 2019, Ap
53	20	1.7	32768	53	US-60-230-446-751	Sequence 751, App	C 126	19	1.6	4417	28	US-09-710-286-3349	Sequence 3349, Ap
54	20	1.7	32768	55	US-60-234-446-423	Sequence 423, App	C 127	19	1.6	4417	28	US-09-717-350-4438	Sequence 4438, Ap
55	20	1.7	32768	55	US-60-234-446-423	Sequence 423, App	C 128	19	1.6	4417	28	US-09-717-350-4438	Sequence 4438, Ap
56	19	1.6	202	13	US-08-917-045-2919	Sequence 2919, Ap	C 129	19	1.6	4417	29	US-09-726-171-2420	Sequence 2420, Ap
57	19	1.6	202	21	US-08-540-229-91915	Sequence 91915, A	C 130	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
58	19	1.6	202	34	US-60-028-710-1175	Sequence 1175, Ap	C 131	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
59	19	1.6	202	35	US-60-032-838-1175	Sequence 839, App	C 132	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
60	19	1.6	207	8	US-08-435-761-839	Sequence 7587, A	C 133	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
61	19	1.6	252	20	US-09-532-315-27587	Sequence 27587, A	C 134	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
62	19	1.6	252	20	US-09-532-315-27587	Sequence 27587, A	C 135	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
63	19	1.6	252	20	US-09-539-806-42459	Sequence 42459, A	C 136	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
64	19	1.6	252	46	US-60-140-769-59	Sequence 59, Appl	C 137	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
65	19	1.6	253	13	US-09-266-620-3901	Sequence 3901, Ap	C 138	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
66	19	1.6	253	16	US-08-978-620-5877	Sequence 5877, Ap	C 139	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
67	19	1.6	263	20	US-09-533-315-27599	Sequence 27599, A	C 140	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
68	19	1.6	282	44	US-60-127-461-6141	Sequence 6141, Ap	C 141	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
69	19	1.6	282	11	US-08-766-606-116	Sequence 116, App	C 142	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
70	19	1.6	282	13	US-08-978-834-116	Sequence 27577, A	C 143	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
71	19	1.6	328	16	US-09-532-315-27577	Sequence 3258, A	C 144	19	1.6	4417	29	US-09-726-787-2641	Sequence 2641, Ap
72	19	1.6	328	16	US-09-205-070-38258	Sequence 38258, A	C 145	18	1.6	153	49	US-09-565-240-53157	Sequence 53157, A
73	19	1.6	328	17	US-09-321-214-29884	Sequence 29884, A	C 146	18	1.6	153	49	US-09-565-240-53157	Sequence 53157, A
74	19	1.6	328	17	US-09-340-633-38258	Sequence 38258, A	C 147	18	1.6	153	49	US-09-565-240-53157	Sequence 53157, A
75													

c 194	18	1.6	568	28	US-09-705-926-7311	Sequence 7311, Ap	267	18	1.6	14280	24	US-09-620-392-10818	Sequence 10818, A
c 195	18	1.6	585	44	US-60-125-818-421	Sequence 421, App	268	18	1.6	14280	24	US-09-702-134-16692	Sequence 16692, A
c 196	18	1.6	593	52	US-60-207-458-6733	Sequence 6733, Ap	c 269	18	1.6	16042	24	US-09-620-392-12451	Sequence 12451, A
c 197	18	1.6	598	17	US-09-354-899-8009	Sequence 8009, Ap	c 270	18	1.6	16042	24	US-09-702-134-26835	Sequence 26835, A
c 198	18	1.6	607	17	US-09-354-899-4447	Sequence 4447, Ap	c 271	18	1.6	17716	35	US-60-038-081-375	Sequence 375, App
c 199	18	1.6	627	49	US-60-172-519-4646	Sequence 4646, Ap	c 272	18	1.6	17716	36	US-60-046-933-385	Sequence 385, App
c 200	18	1.6	692	54	US-60-226-326-4152	Sequence 4152, Ap	c 273	18	1.6	17716	38	US-60-068-228-388	Sequence 388, App
c 201	18	1.6	714	17	US-09-399-720-12272	Sequence 12272, A	c 274	18	1.6	18966	28	US-09-702-134-2311	Sequence 2311, Ap
c 202	18	1.6	715	17	US-09-399-720-12272	Sequence 12272, A	c 275	18	1.6	18966	28	US-09-620-392-25724	Sequence 25724, A
c 203	18	1.6	738	24	US-09-620-608-1354	Sequence 1354, Ap	c 276	18	1.6	20716	53	US-60-213-847-152	Sequence 152, App
c 204	18	1.6	738	46	US-60-144-883-1354	Sequence 1354, Ap	c 277	18	1.6	20716	53	US-60-207-583-168	Sequence 168, App
c 205	18	1.6	784	11	US-08-781-986A-537	Sequence 537, App	c 278	18	1.6	23222	52	PCT-US01-01354-31985	Sequence 31985, A
c 206	18	1.6	784	13	US-08-956-171-537	Sequence 537, App	c 279	18	1.6	23222	52	US-60-207-583-283	Sequence 283, App
c 207	18	1.6	899	55	US-08-956-171B-537	Sequence 537, App	c 280	18	1.6	32768	53	US-60-212-656-4	Sequence 4, Appl1
c 208	18	1.6	1131	50	US-60-236-804-729	Sequence 729, App	c 281	18	1.6	32768	55	US-60-230-435-382	Sequence 382, App
c 209	18	1.6	1134	25	US-09-654-617-1305	Sequence 1305, Ap	c 282	18	1.6	32768	55	US-60-230-445-992	Sequence 992, App
c 210	18	1.6	1134	25	US-09-654-617-1305	Sequence 12278, A	c 283	18	1.6	38023	12	US-08-831-156A-87	Sequence 87, Appl
c 211	18	1.6	1260	1	US-09-684-016-1905	Sequence 1905, Ap	c 284	18	1.6	38023	12	US-08-831-156A-88	Sequence 88, Appl
c 212	18	1.6	1260	1	PCT-US00-08016-1	Sequence 1, Appl1	c 285	18	1.6	40657	48	US-60-167-324-1488	Sequence 1488, Ap
c 213	18	1.6	1260	16	US-09-285-516-1	Sequence 1, Appl1	c 286	18	1.6	40657	48	US-60-167-324-1488	Sequence 1488, Ap
c 214	18	1.6	1266	45	US-60-138-103-25058	Sequence 25058, A	c 287	18	1.6	40958	49	US-09-702-134-25661	Sequence 25661, A
c 215	18	1.6	1281	12	US-08-827-356-561	Sequence 561, App	c 288	18	1.6	40958	49	US-60-173-386-1430	Sequence 1430, Ap
c 216	18	1.6	1281	14	US-09-035-744-1	Sequence 1, Appl1	c 289	18	1.6	40958	49	US-60-175-871-1610	Sequence 1610, Ap
c 217	18	1.6	1392	11	US-09-611-529-2482	Sequence 1, Appl1	c 290	18	1.6	40958	50	US-60-184-775-1470	Sequence 1470, Ap
c 218	18	1.6	1392	11	PCT-US01-01347-177	Sequence 2482, Ap	c 291	18	1.6	40962	51	US-60-191-637-41764	Sequence 41764, A
c 219	18	1.6	1392	11	US-08-785-047-1	Sequence 1, Appl1	c 292	18	1.6	40962	51	US-60-191-700-1576	Sequence 1576, Ap
c 220	18	1.6	1392	11	US-08-785-047A-1	Sequence 1, Appl1	c 293	18	1.6	41094	24	US-60-191-700-1576	Sequence 1576, Ap
c 221	18	1.6	1404	12	US-08-827-356-930	Sequence 1, Appl1	c 294	18	1.6	45921	56	US-60-245-228-30	Sequence 30, Appl
c 222	18	1.6	1404	12	US-08-827-356-930	Sequence 930, App	c 295	18	1.6	45921	57	US-60-258-272-14	Sequence 14, Appl
c 223	18	1.6	1673	23	US-09-611-529-825	Sequence 825, App	c 296	18	1.6	61010	56	US-60-236-804-15	Sequence 15, Appl
c 224	18	1.6	1673	23	US-09-471-275-7068	Sequence 7068, Ap	c 297	18	1.6	61010	56	US-60-243-468-685	Sequence 685, App
c 225	18	1.6	1679	23	US-09-607-200-6698	Sequence 6698, Ap	c 298	18	1.6	68122	19	US-09-528-237A-1055	Sequence 269, App
c 226	18	1.6	1745	18	US-09-457-877-112	Sequence 112, App	c 299	18	1.6	79375	20	US-09-534-859-574	Sequence 1055, Ap
c 227	18	1.6	1843	28	US-09-702-134-19051	Sequence 19051, A	c 300	18	1.6	90425	20	US-09-534-859-574	Sequence 574, App
c 228	18	1.6	1908	24	US-09-620-392-62559	Sequence 62559, A	c 301	18	1.6	90425	20	US-09-534-859-574	Sequence 574, App
c 229	18	1.6	2114	44	US-60-126-248-532	Sequence 532, App	c 302	18	1.6	90425	20	US-09-534-859-574	Sequence 574, App
c 230	18	1.6	2114	44	US-60-126-248-532	Sequence 532, App	c 303	18	1.6	90425	20	US-09-534-859-574	Sequence 574, App
c 231	18	1.6	2114	44	US-60-126-248-532	Sequence 532, App	c 304	18	1.6	90425	20	US-09-534-859-574	Sequence 574, App
c 232	18	1.6	2162	49	US-60-172-373-21742	Sequence 21742, A	c 305	18	1.6	100815	27	US-09-592-412-82	Sequence 82, Appl
c 233	18	1.6	3212	17	US-09-404-520-12569	Sequence 12569, A	c 306	18	1.6	103388	19	US-09-528-237A-1041	Sequence 1041, Ap
c 234	18	1.6	3212	17	US-09-338-425-2523	Sequence 2523, Ap	c 307	18	1.6	103388	19	US-09-528-237A-1041	Sequence 1041, Ap
c 235	18	1.6	3212	25	US-09-644-868-10046	Sequence 10046, A	c 308	18	1.6	110762	23	US-09-611-529-7448	Sequence 7448, Ap
c 236	18	1.6	3212	29	US-09-698-010-15328	Sequence 15328, A	c 309	18	1.6	122075	58	US-60-261-976-31	Sequence 31, Appl
c 237	18	1.6	3212	29	US-09-726-789-4559	Sequence 4559, Ap	c 310	18	1.6	135704	56	US-09-528-237A-1556	Sequence 149, App
c 238	18	1.6	3212	29	US-09-726-789-4559	Sequence 2523, Ap	c 311	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 239	18	1.6	3373	30	US-09-726-789-4559	Sequence 2523, Ap	c 312	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 240	18	1.6	3373	30	US-09-726-789-4559	Sequence 2523, Ap	c 313	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 241	18	1.6	3373	30	US-09-726-789-4559	Sequence 2523, Ap	c 314	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 242	18	1.6	3373	30	US-09-726-789-4559	Sequence 2523, Ap	c 315	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 243	18	1.6	3399	52	US-09-770-173-2697	Sequence 1157, Ap	c 316	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 244	18	1.6	3399	52	US-09-770-173-2697	Sequence 1157, Ap	c 317	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 245	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 318	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 246	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 319	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 247	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 320	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 248	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 321	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 249	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 322	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 250	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 323	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 251	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 324	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 252	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 325	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 253	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 326	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 254	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 327	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 255	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 328	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 256	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 329	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 257	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 330	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 258	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 331	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 259	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 332	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 260	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 333	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 261	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 334	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 262	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 335	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 263	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 336	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 264	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 337	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 265	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 338	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap
c 266	18	1.6	4653	11	US-09-770-173-2697	Sequence 2697, Ap	c 339	17	1.5	153704	56	US-09-528-237A-1556	Sequence 1556, Ap

us-09-151-409-15.oligo.rnppm

Tue May 15 07:26:54 2001

340	17	1.5	239	17	US-09-304-517A-161586	Sequence 161586,	413	17	1.5	322	35	US-60-036-549-2160	Sequence 2160, Ap
341	17	1.5	240	17	US-09-304-517A-205485	Sequence 205485,	414	17	1.5	323	53	US-60-211-750-130772	Sequence 60772, A
342	17	1.5	241	17	US-09-263-191-23026	Sequence 23026, A	c 415	17	1.5	324	53	US-60-217-080-13302	Sequence 13302, A
343	17	1.5	242	17	US-08-917-045-5133	Sequence 5133, Ap	c 416	17	1.5	331	17	US-09-394-745-34427	Sequence 34427, A
344	17	1.5	243	17	US-09-304-517A-153598	Sequence 153598,	c 417	17	1.5	331	19	US-09-521-640-96057	Sequence 96057, A
345	17	1.5	242	17	US-09-534-843-11142	Sequence 11142, A	c 418	17	1.5	331	22	US-09-565-306-54992	Sequence 54992, A
346	17	1.5	242	34	US-60-028-710-33569	Sequence 33569, Ap	c 419	17	1.5	331	25	US-09-684-016-326058	Sequence 326058,
347	17	1.5	242	35	US-60-032-838-3569	Sequence 3569, Ap	c 420	17	1.5	331	25	US-09-684-016-326058	Sequence 326058,
348	17	1.5	247	16	US-09-081-109-1377	Sequence 1377, Ap	c 421	17	1.5	339	16	US-09-283-466-2023	Sequence 2023, Ap
349	17	1.5	247	16	US-09-244-000A-57733	Sequence 57733, A	c 422	17	1.5	341	17	US-09-304-517A-281337	Sequence 281337,
350	17	1.5	247	16	US-09-263-191-30715	Sequence 30715, A	c 423	17	1.5	341	53	US-60-211-750-33492	Sequence 33492, A
351	17	1.5	247	16	US-09-283-466-39490	Sequence 39490, A	c 424	17	1.5	343	16	US-09-270-767-12437	Sequence 12437, A
352	17	1.5	247	21	US-09-540-764-43006	Sequence 43006, A	c 425	17	1.5	349	17	US-09-304-517A-276368	Sequence 276368,
353	17	1.5	247	36	US-09-048-727-1377	Sequence 1377, Ap	c 426	17	1.5	349	18	US-09-440-687-29614	Sequence 29614, A
354	17	1.5	248	17	US-09-304-517A-200417	Sequence 200417,	c 427	17	1.5	349	46	US-60-144-084-17377	Sequence 17377, A
355	17	1.5	249	17	US-09-304-517A-146411	Sequence 146411,	c 428	17	1.5	352	11	US-08-766-606-370	Sequence 370, App
356	17	1.5	249	22	US-09-552-086-19319	Sequence 19319, A	c 429	17	1.5	352	13	US-08-979-854A-370	Sequence 27578, A
357	17	1.5	249	45	US-09-565-240-31174	Sequence 31174, A	c 430	17	1.5	352	20	US-09-532-315-27578	Sequence 40790, A
358	17	1.5	249	45	US-09-133-691-5297	Sequence 5297, Ap	c 431	17	1.5	353	53	US-60-211-750-40790	Sequence 395805,
359	17	1.5	250	21	US-09-304-517A-199449	Sequence 199449,	c 432	17	1.5	353	25	US-09-654-617-395805	Sequence 395805,
360	17	1.5	250	21	US-09-540-229-117439	Sequence 117439,	c 433	17	1.5	356	22	US-09-684-016-395805	Sequence 29883,
361	17	1.5	252	16	US-09-263-191-15479	Sequence 15479, A	c 434	17	1.5	361	22	US-09-565-240-29883	Sequence 3837, Ap
362	17	1.5	252	16	US-09-294-093A-3378	Sequence 3378, A	c 435	17	1.5	361	45	US-60-133-691-3837	Sequence 1987, Ap
363	17	1.5	252	16	US-09-284-093B-3849	Sequence 3849, Ap	c 436	17	1.5	361	50	US-60-189-657-1987	Sequence 675, App
364	17	1.5	252	40	US-09-082-567-3849	Sequence 3849, Ap	c 437	17	1.5	363	14	US-09-016-434-675	Sequence 11143, A
365	17	1.5	254	17	US-09-304-517A-158260	Sequence 158260,	c 438	17	1.5	363	20	US-09-534-843-11143	Sequence 8959, Ap
366	17	1.5	255	17	US-09-306-349-8390	Sequence 8390, Ap	c 439	17	1.5	364	18	US-09-431-517-8959	Sequence 93842, A
367	17	1.5	255	17	US-09-244-000A-42533	Sequence 42533, A	c 440	17	1.5	365	19	US-09-521-640-93842	Sequence 42177, A
368	17	1.5	260	13	US-08-917-045-2658	Sequence 2658, Ap	c 441	17	1.5	367	22	US-09-565-240-42177	Sequence 10471, A
369	17	1.5	260	21	US-09-540-229-73215	Sequence 73215, A	c 442	17	1.5	367	23	US-09-606-977-10471	Sequence 3542, Ap
370	17	1.5	260	34	US-60-038-710-914	Sequence 914, App	c 443	17	1.5	367	45	US-60-133-692-3542	Sequence 19652, A
371	17	1.5	261	17	US-09-032-838-914	Sequence 914, App	c 444	17	1.5	368	19	US-09-515-128-19652	Sequence 30194, A
372	17	1.5	261	17	US-09-304-517A-189371	Sequence 189371,	c 445	17	1.5	368	22	US-09-565-240-30194	Sequence 8337, Ap
373	17	1.5	261	21	US-09-540-229-181527	Sequence 181527,	c 446	17	1.5	368	45	US-09-577-408-8337	Sequence 4210, Ap
374	17	1.5	261	40	US-60-089-037-425	Sequence 425, App	c 447	17	1.5	368	20	US-60-133-691-4210	Sequence 2700, Ap
375	17	1.5	262	11	US-08-733-814-275	Sequence 275, App	c 448	17	1.5	371	26	US-09-663-779-2700	Sequence 11402, A
376	17	1.5	262	17	US-09-304-517A-183720	Sequence 183720,	c 449	17	1.5	372	16	US-09-270-767-11402	Sequence 51800, A
377	17	1.5	262	17	US-09-333-535A-9682	Sequence 9682, Ap	c 450	17	1.5	373	53	US-60-207-458-51800	Sequence 13492, A
378	17	1.5	262	21	US-09-540-499-25360	Sequence 25360, A	c 451	17	1.5	375	53	US-60-217-080-13492	Sequence 97830, A
379	17	1.5	262	21	US-09-540-764-27327	Sequence 27327, A	c 452	17	1.5	376	19	US-09-521-640-97830	Sequence 193783, A
380	17	1.5	262	22	US-09-552-086-23067	Sequence 23067, A	c 453	17	1.5	378	19	US-09-521-640-193783	Sequence 20641, A
381	17	1.5	262	42	US-60-108-390-125	Sequence 125, App	c 454	17	1.5	378	46	US-60-140-769-20641	Sequence 6146, Ap
382	17	1.5	265	16	US-09-263-191-25425	Sequence 25425, A	c 455	17	1.5	380	20	US-60-202-213-6146	Sequence 16366, A
383	17	1.5	265	17	US-09-304-517A-156611	Sequence 156611,	c 456	17	1.5	380	25	US-09-534-843-16366	Sequence 459086,
384	17	1.5	265	17	US-09-304-517A-162628	Sequence 162628,	c 457	17	1.5	384	25	US-09-654-617-459086	Sequence 459086,
385	17	1.5	265	20	US-09-539-806-47440	Sequence 47440, A	c 458	17	1.5	384	23	US-09-684-016-459086	Sequence 8301, Ap
386	17	1.5	266	16	US-09-244-000A-5512	Sequence 5512, Ap	c 459	17	1.5	385	23	US-09-606-977-8301	Sequence 22153, A
387	17	1.5	266	16	US-09-244-000A-25623	Sequence 25623, A	c 460	17	1.5	387	19	US-09-505-532-22153	Sequence 8764, Ap
388	17	1.5	266	16	US-09-304-517A-285706	Sequence 285706,	c 461	17	1.5	387	19	US-09-521-640-8764	Sequence 23801, A
389	17	1.5	272	15	US-09-199-129A-4159	Sequence 4159, Ap	c 462	17	1.5	391	17	US-09-362-510-23801	Sequence 23801, A
390	17	1.5	272	15	US-09-304-517A-278709	Sequence 278709,	c 463	17	1.5	391	17	US-09-362-510A-23801	Sequence 176172, A
391	17	1.5	273	16	US-09-283-466-34185	Sequence 34185, A	c 464	17	1.5	392	19	US-09-521-640-176172	Sequence 40964, A
392	17	1.5	273	52	US-60-209-830-42825	Sequence 42825, A	c 465	17	1.5	393	51	US-60-197-873-40964	Sequence 2862, Ap
393	17	1.5	274	16	US-09-244-000A-63998	Sequence 63998, A	c 466	17	1.5	397	6	US-08-220-691-2862	Sequence 2862, Ap
394	17	1.5	274	17	US-09-304-517A-294632	Sequence 294632,	c 467	17	1.5	397	23	US-08-220-691B-2862	Sequence 26342, A
395	17	1.5	274	22	US-09-552-086-4321	Sequence 4321, Ap	c 468	17	1.5	399	19	US-09-606-977-26342	Sequence 67394, A
396	17	1.5	276	15	US-09-199-129A-2388	Sequence 2388, Ap	c 469	17	1.5	399	23	US-09-528-409-67394	Sequence 29844, A
397	17	1.5	276	16	US-09-215-179A-4949	Sequence 4949, Ap	c 470	17	1.5	400	18	US-09-421-106-29844	Sequence 29857, A
398	17	1.5	278	16	US-09-244-000A-98249	Sequence 98249, A	c 471	17	1.5	400	19	US-09-521-640-29857	Sequence 29857, A
399	17	1.5	281	52	US-60-207-458-21127	Sequence 21127, A	c 472	17	1.5	401	24	US-09-637-086A-38051	Sequence 38051, Ap
400	17	1.5	282	15	US-09-199-129A-1731	Sequence 1731, Ap	c 473	17	1.5	405	16	US-09-248-797-3903	Sequence 3903, Ap
401	17	1.5	282	21	US-09-540-499-20865	Sequence 20865, A	c 474	17	1.5	405	17	US-09-346-956-76	Sequence 76, App1
402	17	1.5	283	11	US-08-798-074-7560	Sequence 7560, Ap	c 475	17	1.5	405	22	US-09-565-240-53016	Sequence 53016, A
403	17	1.5	285	16	US-09-244-000A-19723	Sequence 19723, A	c 476	17	1.5	408	19	US-09-528-409-82166	Sequence 82166, A
404	17	1.5	285	17	US-09-304-517A-289315	Sequence 289315,	c 477	17	1.5	408	22	US-09-565-240-29718	Sequence 29718, A
405	17	1.5	287	17	US-09-306-349-39619	Sequence 39619, A	c 478	17	1.5	408	45	US-60-133-691-3655	Sequence 3655, Ap
406	17	1.5	288	17	US-09-304-517A-188395	Sequence 188395,	c 479	17	1.5	409	22	US-09-565-306-29369	Sequence 10963, A
407	17	1.5	291	20	US-09-534-843-16363	Sequence 16363, A	c 480	17	1.5	409	25	US-09-654-617-292009	Sequence 292009,
408	17	1.5	306	14	US-09-070-694-3294	Sequence 3294, Ap	c 481	17	1.5	409	27	US-09-684-016-292009	Sequence 292009,
409	17	1.5	317	25	US-09-654-617-26497	Sequence 26497, A	c 482	17	1.5	410	19	US-09-521-640-228973	Sequence 228973,
410	17	1.5	317	27	US-09-684-016-26497	Sequence 26497, A	c 483	17	1.5	411	50	US-60-189-657-3549	Sequence 3549, Ap
411	17	1.5	319	21	US-09-540-499-4660	Sequence 4660, Ap	c 484	17	1.5	416	52	US-60-207-458-33443	Sequence 33443, A
412	17	1.5	322	14	US-09-036-591-2160	Sequence 2160, Ap	c 485	17	1.5				

486	17	1.5	US-09-790-478-5702	Sequence 5702, Ap	c 559	17	1.5	494	23	US-09-605-701-8206
487	17	1.5	US-09-790-478-5702	Sequence 12275, A	c 560	17	1.5	494	25	US-09-654-617-133885
488	17	1.5	US-09-528-409-449	Sequence 449, App	c 561	17	1.5	494	27	US-09-684-016-133885
489	17	1.5	US-09-362-510-109	Sequence 109, App	c 562	17	1.5	494	29	US-09-684-016-133885
490	17	1.5	US-09-362-510A-109	Sequence 109, App	c 563	17	1.5	501	16	US-09-248-796-5416
491	17	1.5	US-09-521-640-98022	Sequence 98022, A	c 564	17	1.5	501	18	US-09-609-409-5416
492	17	1.5	US-09-421-106-10175	Sequence 10175, A	c 565	17	1.5	502	16	US-09-441-548A-750
493	17	1.5	US-09-521-640-65352	Sequence 65352, A	c 566	17	1.5	506	6	US-08-220-691-4782
494	17	1.5	US-09-521-640-65352	Sequence 65352, A	c 567	17	1.5	506	6	US-08-220-691-4782
495	17	1.5	US-09-521-640-254329	Sequence 254329, A	c 568	17	1.5	506	25	US-09-654-617-16809
496	17	1.5	US-09-540-229-191382	Sequence 191382, A	c 569	17	1.5	506	27	US-09-684-016-16809
497	17	1.5	US-09-521-640-66182	Sequence 66182, A	c 570	17	1.5	509	19	US-09-505-532-40043
498	17	1.5	US-09-404-284-1226	Sequence 1226, Ap	c 571	17	1.5	519	25	US-09-654-617-192978
499	17	1.5	US-09-524-038-1226	Sequence 1226, Ap	c 572	17	1.5	519	27	US-09-684-016-192978
500	17	1.5	US-09-078-617-2979	Sequence 2979, Ap	c 573	17	1.5	522	20	US-09-534-843-15370
501	17	1.5	US-09-248-797-931	Sequence 931, App	c 574	17	1.5	525	28	US-09-716-972-4353
502	17	1.5	US-09-399-720-2180	Sequence 2180, Ap	c 575	17	1.5	528	13	US-08-993-002A-2534
503	17	1.5	US-09-505-532-31362	Sequence 31362, A	c 576	17	1.5	531	13	US-09-304-517A-216115
504	17	1.5	US-06-155-006-3380	Sequence 3380, Ap	c 577	17	1.5	531	22	US-09-552-086-13069
505	17	1.5	US-09-540-234-8917	Sequence 8917, Ap	c 578	17	1.5	538	51	US-06-196-868-15144
506	17	1.5	US-09-654-617-447409	Sequence 447409, A	c 579	17	1.5	539	52	US-60-207-458-7333
507	17	1.5	US-09-684-016-447409	Sequence 447409, A	c 580	17	1.5	539	52	US-60-207-458-7333
508	17	1.5	US-09-699-998-661	Sequence 661, App	c 581	17	1.5	540	22	US-09-563-309A-13421
509	17	1.5	US-08-401-791A-14075	Sequence 14075, A	c 582	17	1.5	541	20	US-09-563-309A-13421
510	17	1.5	US-08-401-791B-14075	Sequence 14075, A	c 583	17	1.5	541	52	US-09-534-843-16368
511	17	1.5	US-08-993-002A-2533	Sequence 2533, Ap	c 584	17	1.5	544	22	US-60-202-214-3311
512	17	1.5	US-08-976-343-2716	Sequence 2716, Ap	c 585	17	1.5	544	22	US-09-563-309A-16192
513	17	1.5	US-08-976-343A-2716	Sequence 2716, Ap	c 586	17	1.5	550	19	US-09-505-532-9752
514	17	1.5	US-06-197-872-3544	Sequence 3544, A	c 587	17	1.5	552	28	US-09-702-134-35619
515	17	1.5	US-09-521-640-83799	Sequence 83799, A	c 588	17	1.5	554	44	US-09-304-517A-210707
516	17	1.5	US-09-528-409-68609	Sequence 68609, A	c 589	17	1.5	555	23	US-60-125-818-10688
517	17	1.5	US-60-211-750-40366	Sequence 40366, A	c 590	17	1.5	555	25	US-09-607-200-1684
518	17	1.5	US-09-399-720-13626	Sequence 13626, A	c 591	17	1.5	557		


```
QY 481 ccatggagagtcggttatactagcaccagagataagctaaataattattgac 540
Db 481 ccatggagagtcggttatactagcaccagagataagctaaataattattgac 540
QY 541 gatcttgacgcaaaagaaacatgaattacaaacacagaatgagaagtattctctgcagaac 600
Db 541 gatcttgacgcaaaagaaacatgaattacaaacacagaatgagaagtattctctgcagaac 600
QY 601 ggtgatggttaactctaggaagattatagaagattcttgacgcaaaacaaatcccgcaatacaa 660
Db 601 ggtgatggttaactctaggaagattatagaagattcttgacgcaaaacaaatcccgcaatacaa 660
QY 661 aatatacgtttacgtacgcaaaacaaaggaacttaaaagcagattagagaatgcaatggaa 720
Db 661 aatatacgtttacgtacgcaaaacaaaggaacttaaaagcagattagagaatgcaatggaa 720
QY 721 gttcaggaagagatttttaagagagctggtaacctgttgatcaggtttacacaattatat 780
Db 721 gttcaggaagagatttttaagagagctggtaacctgttgatcaggtttacacaattatat 780
QY 781 actaaacataatagtaattaccacaacataatgcacaaagctggcagacttgacctgaga 840
Db 781 actaaacataatagtaattaccacaacataatgcacaaagctggcagacttgacctgaga 840
QY 841 caaaaggctgaatatctaaaggccttaataatgattggctgagagctgttacaagagtta 900
Db 841 caaaaggctgaatatctaaaggccttaataatgattggctgagagctgttacaagagtta 900
QY 901 aatatcgatgctgcgactagctctcagacagatactctggaagaaagtcacagaacgtgct 960
Db 901 aatatcgatgctgcgactagctctcagacagatactctggaagaaagtcacagaacgtgct 960
QY 961 gacaagtttgatagataaaacaaatcactgtaaaacttaagaatagtgacttaagtttaatt 1020
Db 961 gacaagtttgatagataaaacaaatcactgtaaaacttaagaatagtgacttaagtttaatt 1020
QY 1021 aataaagcgttaaaagatcataatgatgagtttaactgaagagtgagtaagttaagtttaatt 1080
Db 1021 aataaagcgttaaaagatcataatgatgagtttaactgaagagtgagtaagttaagtttaatt 1080
QY 1081 aaactacgtataaaatgataaaatcactactctgaaagaaagctagtaaaatcagaattagag 1140
Db 1081 aaactacgtataaaatgataaaatcactactctgaaagaaagctagtaaaatcagaattagag 1140
QY 1141 gcacgttaagtaaaagcctt 1158
Db 1141 gcacgttaagtaaaagcctt 1158
```

RESULT 2

US-07-945-954A-2

; Sequence 2, Application US/07945954A

; GENERAL INFORMATION:

; APPLICANT: Dale, James B.

; APPLICANT: Lederer, James W.

; TITLE OF INVENTION: Recombinant Multivalent M Protein

; TITLE OF INVENTION: Vaccine

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weiser & Associates

; STREET: 230 South Fifteenth Street, Suite 500

; CITY: Philadelphia

; STATE: Pennsylvania

; COUNTRY: U.S.A.

; ZIP: 19102

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/945,954A

; FILING DATE: 16-SEP-1992

```
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 741 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
US-07-945-954A-2
```

Query Match 34.7%; Score 402; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 8.9e-197;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggtccgcgcgtgactaggggtacaataaatgacccgcaaaagagcaaaagaagctcttgac 249

Db 340 GGATCCGCGGTGACTAGGGGTACAATAATGACCCCAAGAGCAAAAGCTCTTGAC 399

QY 250 aatgatgagctagaacacccatgacttaaaactaagaatgaagggttaaaactgagaat 309

Db 400 AAGTATGACGTAGAAAACCATGACTTTAAAACTAAGAATCAAGGGTTAAAACTGAGAA 459

QY 310 gaagggttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaactgag 369

Db 460 GAAGGGTTAAACCTGAGAACTGAAGGGTTAAAACTGAGAACTGAAGGGTTAAAACTGAG 519

QY 370 gtgcagagagtttctctagggggacggttagaaaacccgggacaaagcagcagaactctt 429

Db 520 GTCGACAGAGTGTTCCTAGGGGACGCTAGAAAACCCGGACAAAGCAGCAGAACTTCTT 579

QY 430 aacaagtatgacgtagagaactctatgtttacaagctataatgacaagttaccatggaga 489

Db 580 AACAGTATGACGTAGAGAACTCTATGTTACAAAGCTAATAATGACAAGTTACCATGGAGA 639

QY 490 gtgcgttatactagcgacacgcagagaataagctaaataaaataattatgacgactctgac 591

Db 640 GTGCGTTATCTAGGCATAGCCGACAGAGTAAGCTAAAAAATAATTTATTGACGATCTTGAC 699

QY 550 qcaaaagaacatgaattacaacaacagaaatgagaagttatct 591

Db 700 GCAAAAGAACATGAATTACAACACAGAAATGAGAAGTTATCT 741

RESULT 3

US-07-945-954A-1

; Sequence 1, Application US/07945954A

; GENERAL INFORMATION:

; APPLICANT: Dale, James B.

; APPLICANT: Lederer, James W.

; TITLE OF INVENTION: Recombinant Multivalent M Protein

; TITLE OF INVENTION: Vaccine

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weiser & Associates

; STREET: 230 South Fifteenth Street, Suite 500

; CITY: Philadelphia

; STATE: Pennsylvania

; COUNTRY: U.S.A.

ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,954A
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372.5672P
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 861 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..861
US-07-945-954A-1

Query Match 20.7%; Score 240; DB 3; Length 861;
Best Local Similarity 100.0%; Pred. No. 5.6e-113;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtgcgactaggtctcagacagatactctggaaaaagtcacaagaacgtgctgacaagt 969
Db 4 GTCCGACTAGGTCTCAGACAGATACCTCTGGAAAAGTACAGACGCTGCTGACAAGTTT 63
QY 970 gagatgaaaaacaatacgtttaaaacttaagaatagtgacttaagttttaataaagcg 1029
Db 64 GAGATAGAAAACAATAACGTTAAACCTTAAGAATAGTACTTAAGTTTAAATAAAGCG 123
QY 1030 ttaaaagatcataatgatgagtttaactgaagattgagtaagtctaaagagaaactacgt 1089
Db 124 TTAAGAATCATATGATGAGTTTAACTGAAGAGTTGAGTAATGCTTAAGAGAAACTACGT 183
QY 1090 aaaaatgataatcactatctgaaaagctagtaaaattcaagaattagaggcagtaag 1149
Db 184 AAAATGATAATCACTATCTGAAAAGCTAGTAAATTCAGAATTAGAGGCAGTAAG 243

RESULT 4
US-07-945-954A-7
Sequence 7, Application US/07945954A
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: Recombinant Multivalent M Protein
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,954A
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372.5672P
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 522 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..519
US-07-945-954A-7

Query Match 18.7%; Score 217; DB 3; Length 522;
Best Local Similarity 100.0%; Pred. No. 4.4e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtgcgactaggtctcagacagatactctggaaaaagtcacaagaacgtgctgacaagt 969
Db 4 GTCCGACTAGGTCTCAGACAGATACCTCTGGAAAAGTACAGACGCTGCTGACAAGTTT 63
QY 970 gagatgaaaaacaatacgtttaaaacttaagaatagtgacttaagttaataaagcg 1029
Db 64 GAGATAGAAAACAATAACGTTAAACCTTAAGAATAGTACTTAAGTTTAAATAAAGCG 123
QY 1030 ttaaaagatcataatgatgagtttaactgaagattgagtaagtctaaagagaaactacgt 1089
Db 124 TTAAGAATCATATGATGAGTTTAACTGAAGAGTTGAGTAATGCTTAAGAGAAACTACGT 183
QY 1090 aaaaatgataatcactatctgaaaagctagtaaaattcaagaattagaggcagtaag 1126
Db 184 AAAATGATAATCACTATCTGAAAAGCTAGTAAATAAGCTAGTAAAA 220

RESULT 5
US-07-945-954A-6
Sequence 6, Application US/07945954A
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: Recombinant Multivalent M Protein
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/945,954A
;; FILING DATE: 16-SEP-1992
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weiser, Gerard J.
;; REGISTRATION NUMBER: 19,763
;; REFERENCE/DOCKET NUMBER: 372.5672P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-875-8383
;; TELEFAX: 215-875-8394
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 741 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Streptococcus pyogenes
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..741
US-07-945-954A-6

Query Match 18.7%; Score 217; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 4.4e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggctcagacagatactctggaaaaagtagcaaacgctgctgacaagttt 969
|||||
Db 406 GTCGCGACTAGGCTCAGACAGATACCTCGAAAAAGTACAAAGACGCTGTCACAGTTT 465

QY 970 gagatagaaaaacaatacgttaaaacttaagaataagtagtaagttttaataataaagcg 1029
|||||
Db 466 GAGATAGAAAACAATAACGTTAAACCTTAAGAATAGTGACTTAAAGTTTAAATAATAAGCG 525

QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagagaaactacgt 1089
|||||
Db 526 TTAAGAAGATCAATGATGAGTTAACTGAAGAGTTTGAGTAATCTTAAGTAATAAGAACTACGT 585

QY 1090 aaaaatgataaatacactatctgaaaaagcctagtaaaa 1126
|||||
Db 586 AAAAATGATAAATCACTATCTGAAAAAAGCTAGTAAAA 622

RESULT 6
US-07-945-954A-3
; Sequence 3, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992

;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weiser, Gerard J.
;; REGISTRATION NUMBER: 19,763
;; REFERENCE/DOCKET NUMBER: 372.5672P
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-875-8383
;; TELEFAX: 215-875-8394
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 822 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Streptococcus pyogenes
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..822
US-07-945-954A-3

Query Match 18.7%; Score 217; DB 3; Length 822;
Best Local Similarity 100.0%; Pred. No. 4.4e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggctcagacagatactctggaaaaagtagcaaacgctgctgacaagttt 969
|||||
Db 4 GTCGCGACTAGGCTCAGACAGATACCTCGAAAAAGTACAAAGACGCTGTCACAGTTT 63

QY 970 gagatagaaaaacaatacgttaaaacttaagaataagtagtaagttttaataataaagcg 1029
|||||
Db 64 GAGATAGAAAACAATAACGTTAAACCTTAAGAATAGTGACTTAAAGTTTAAATAATAAGCG 123

QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagagaaactacgt 1089
|||||
Db 124 TTAAGAAGATCAATGATGAGTTAACTGAAGAGTTTGAGTAATCTTAAGTAATAAGAACTACGT 183

QY 1090 aaaaatgataaatacactatctgaaaaagcctagtaaaa 1126
|||||
Db 184 AAAAATGATAAATCACTATCTGAAAAAAGCTAGTAAAA 220

RESULT 7
US-07-945-954A-8
; Sequence 8, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.

```
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1029
US-07-945-954A-8

Query Match          18.7%; Score 217; DB 3; Length 1029;
Best Local Similarity 100.0%; Pred. No. 4.4e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatctctggaagaaagtacaagacgtgctgacaagttt 969
DB 406 gtcgcgactaggtctcagacagatctctggaagaaagtacaagacgtgctgacaagttt 465

QY 970 gagatagaaaaacaatacgttataaaacttaagaatagtgacttaagttaataataaagcg 1029
DB 466 GAGATAGAAAAACAATACTGTTAAACTTAAAGAAATAGTGACTTAACTTAAATAATAAAGCG 525

QY 1030 ttaaaagatcataatgatgagtttaactgaagagtggtgagtaagtctaagagaaactcgt 1089
DB 526 TTAAGATGATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 585

QY 1090 aaaaatgataatcactatctctgaaaaagctagtaaaa 1126
DB 586 AAAAATGATAATCACTATCTGAAAAAGCTAGTAAAA 622

RESULT 8
US-08-325-278-5
; Sequence 5, Application US/08325278
; GENERAL INFORMATION:
; APPLICANT: Bivick, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,278
; FILING DATE: 26-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMeesters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 450023.401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
```

```
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1329
US-08-325-278-5

Query Match          13.0%; Score 151; DB 7; Length 1332;
Best Local Similarity 100.0%; Pred. No. 6.1e-67;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 598 aacggtgatgttaactcctagggaaagttatagaagatcttgcagcaaacatccccgcaata 657
DB 1 AACCGTGATGGTGAATCCTAGGGAAGTTATAGAAGATCTTCCAGCAAAACAATCCCGCAATA 60

QY 658 caaaatatacgttttacgttcacgaaaaaacaaggacttaaaagcgagattagagaatgcaatg 717
DB 61 CAAAATATACGTTTACGTCACGAAAAACAAGGACTTAAAAAGCGAGATTAGAGAATGCAATG 120

QY 718 gaagttgcggaagagatatttaagagagctg 748
DB 121 GAAGTTGCAGGAAGAGATTTTAAAGAGAGCTG 151

RESULT 9
US-07-945-954A-4
; Sequence 4, Application US/07945954A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: Recombinant Multivalent M Protein
; TITLE OF INVENTION: Vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```


ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-07-945-954A-4

Query Match 4.6%; Score 53; DB 3; Length 561;
Best Local Similarity 100.0%; Pred. No. 3e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggatccgcctgactagggtacataataatgaccgcgaagcgaagcgaagc 241
|||||
Db 139 GGATCCCGCTGACTAGGGGTACAATAATGACCCGCAAGAGCAAGAGC 191

RESULT 10
US-07-945-954A-9
Sequence 9, Application US/07945954A
GENERAL INFORMATION:

APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: Recombinant Multivalent M Protein
TITLE OF INVENTION: Vaccine
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,954A
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372.5672P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 204 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:

NAME/KEY: CDS
LOCATION: 1..201
US-07-945-954A-9

Query Match 4.5%; Score 52; DB 3; Length 204;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggatccgcctgactagggtacataataatgaccgcgaagcgaagcgaagc 241
|||||
Db 49 GGATCCCGCTGACTAGGGGTACAATAATGACCCGCAAGAGCAAGAGC 100

RESULT 11
US-07-945-954A-10
Sequence 10, Application US/07945954A
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: Recombinant Multivalent M Protein
TITLE OF INVENTION: Vaccine
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,954A
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372.5672P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 408 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..405
US-07-945-954A-10

Query Match 4.5%; Score 52; DB 3; Length 408;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggatccgcctgactagggtacataataatgaccgcgaagcgaagcgaagc 241
|||||
Db 49 GGATCCCGCTGACTAGGGGTACAATAATGACCCGCAAGAGCAAGAGC 100

RESULT 12
US-07-945-954A-5
Sequence 5, Application US/07945954A
GENERAL INFORMATION:

APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: Recombinant Multivalent M Protein
TITLE OF INVENTION: Vaccine
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia

```
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,954A
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5672P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 918 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..915
; US-07-945-954A-5

Query Match 4.5%; Score 52; DB 3; Length 918;
Best Local Similarity 100.0%; Pred. No. 9.9e-16;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggatccgcctgactagggtacataataatgaccgcgaagcaaaagaaag 241
|||||
DB 49 GGATCCCGCTGACTAGGGGTACATAATAATGACCCGCAAGAGCAAAAGAG 100

RESULT 13
US-07-945-860A-2
; Sequence 2, Application US/07945860A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: Antigen of Hybrid M Protein and Carrier
; TITLE OF INVENTION: for Group A Streptococcal Vaccine
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,860A
; FILING DATE: 19920916
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5770P
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 765 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; US-07-945-860A-2

Query Match 4.1%; Score 48; DB 3; Length 765;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 atggtcgcgactaggctcagacagatactctgaaaaagtcacaaagaa 54
|||||
DB 1 ATGCTCGCGACTAGGTCTCAGACAGATACTCTGAAAAAGTACAAGAA 48

RESULT 14
US-07-945-860B-2
; Sequence 2, Application US/07945860B
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; TITLE OF INVENTION: FOR GROUP A STREPTOCOCCAL VACCINE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,860B
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5770P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..762
; US-07-945-860B-2
```

Query Match 4.1%; Score 48; DB 13; Length 765;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 atggtcgcgactaggctctcagacagatactctggaaaaagtacaagaa 54
|||||
Db 1 ATGGTCGCGACTAGGTCTCAGACAGATACTCTGGAAAAAGTACAAGAA 48

RESULT 15
US-08-914-479-3
; Sequence 3, Application US/08914479
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 481112.404C2
; CURRENT APPLICATION NUMBER: US/08/914.479
; CURRENT FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; PRIOR FILING DATE: 1992-09-16
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: An antigen of M5 and a carrier of the
; OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479-3

Query Match 4.1%; Score 48; DB 13; Length 765;
Best Local Similarity 100.0%; Pred. No. 1.2e-13;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 atggtcgcgactaggctctcagacagatactctggaaaaagtacaagaa 54
|||||
Db 1 atggtcgcgactaggctctcagacagatactctggaaaaagtacaagaa 48

Search completed: May 13, 2001, 06:46:59
Job time: 3538 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 05:29:41 ; Search time 86.88 Seconds
(without alignments)
5000.619 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagcatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 262727 seqs, 187588001 residues

Word size : 10

Total number of hits satisfying chosen parameters: 122670

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending_Patents_NA_New.*

1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	1.6	275	5	US-09-540-212A-66707
2	18	1.6	784	4	US-08-956-171C-537
3	18	1.6	4813	1	PCT-US01-0409A-471
4	18	1.6	5623	1	PCT-US01-0409A-2439
5	18	1.6	283061	6	US-60-248-505-134
6	17	1.5	1332	5	US-09-540-212A-13409
7	17	1.5	1682	5	US-09-801-833-7074
8	17	1.5	3929	5	US-09-543-679A-2639
9	17	1.5	9834	4	US-08-956-171C-37
10	17	1.5	12447	5	US-09-543-679A-2640
11	17	1.5	126883	6	US-60-248-505-614
12	17	1.5	924430	5	US-09-335-032-12216
13	16	1.4	95	5	US-09-540-212A-37688
14	16	1.4	120	5	US-09-540-212A-21769
15	16	1.4	165	5	US-09-815-343-1168
16	16	1.4	240	5	US-09-815-343-220
17	16	1.4	245	5	US-09-540-212A-2062
18	16	1.4	274	5	US-09-540-212A-57406
19	16	1.4	266	6	US-60-010-803-3361
20	16	1.4	303	5	US-09-487-566A-3760
21	16	1.4	317	5	US-09-724-866A-19388
22	16	1.4	375	5	US-09-540-212A-61419
23	16	1.4	402	4	US-08-276-1630-5229
24	16	1.4	404	5	US-09-737-223-18462
25	16	1.4	426	1	PCT-US01-01339-992
26	16	1.4	426	1	PCT-US01-01339-992
27	16	1.4	452	5	US-09-737-223-9871

28	16	1.4	453	1	PCT-US01-01339-9895
29	16	1.4	473	5	US-09-737-223-31535
30	16	1.4	474	5	US-09-737-223-16219
31	16	1.4	486	5	US-09-724-866A-6168
32	16	1.4	489	5	US-09-724-866A-5608
33	16	1.4	506	5	US-09-737-223-9337
34	16	1.4	511	5	US-09-827-271-135
35	16	1.4	522	6	US-60-281-593-482
36	16	1.4	564	1	PCT-US01-01339-1755
37	16	1.4	564	1	PCT-US01-01329-715
38	16	1.4	572	5	US-09-801-833-6254
39	16	1.4	647	1	PCT-US01-01339-8336
40	16	1.4	647	1	PCT-US01-01339-8338
41	16	1.4	675	4	US-08-956-171C-887
42	16	1.4	744	4	US-08-956-171C-966
43	16	1.4	1166	5	US-09-603-124B-433
44	16	1.4	1361	5	US-09-569-899-9
45	16	1.4	1599	5	US-09-739-449-5554
46	16	1.4	2114	1	PCT-US01-01302-178
47	16	1.4	2138	1	PCT-US01-01339-8314
48	16	1.4	2180	1	PCT-US01-01302-41
49	16	1.4	2225	1	PCT-US01-01301-359
50	16	1.4	2620	1	PCT-US01-01302-179
51	16	1.4	2663	5	US-09-654-935A-238
52	16	1.4	3167	1	PCT-US00-35017A-357
53	16	1.4	3710	5	US-09-543-679A-2477
54	16	1.4	4842	1	PCT-US01-01339-8504
55	16	1.4	4842	1	PCT-US01-01339-8505
56	16	1.4	4842	1	PCT-US01-01329-3032
57	16	1.4	4842	1	PCT-US01-01329-3033
58	16	1.4	8365	5	US-09-543-679A-3012
59	16	1.4	8414	1	PCT-US01-01312-1278
60	16	1.4	8987	6	US-60-248-505-628
61	16	1.4	11391	1	PCT-US01-01339-9893
62	16	1.4	13086	4	US-08-956-171C-16
63	16	1.4	14273	1	PCT-US01-01339-10125
64	16	1.4	15223	5	US-09-444-067-1
65	16	1.4	23694	6	US-60-248-505-601
66	16	1.4	24034	6	US-60-248-823-40
67	16	1.4	28080	6	US-60-248-505-70
68	16	1.4	32212	1	PCT-US01-01339-8770
69	16	1.4	32212	1	PCT-US01-01329-3299
70	16	1.4	36519	6	US-60-248-505-370
71	16	1.4	39789	6	US-60-248-505-157
72	16	1.4	49722	6	US-60-248-505-501
73	16	1.4	50193	6	US-60-248-505-351
74	16	1.4	50713	6	US-60-248-505-85
75	16	1.4	55193	6	US-60-248-505-511
76	16	1.4	70661	6	US-60-248-505-133
77	16	1.4	82769	6	US-60-248-823-20
78	16	1.4	106174	6	US-60-248-505-419
79	16	1.4	162025	5	US-09-834-700-13
80	16	1.4	162025	5	US-09-834-700-14
81	16	1.4	162025	5	US-09-834-700-17
82	16	1.4	162025	5	US-09-834-700-18
83	16	1.4	237961	6	US-60-251-317-2
84	16	1.4	249597	5	US-60-248-505-172
85	16	1.4	315341	6	US-09-335-032-12206
86	16	1.4	318095	5	US-09-739-449-215
87	16	1.4	334854	6	US-60-248-505-28
88	16	1.4	443925	6	US-60-248-505-299
89	16	1.4	443925	6	US-60-248-505-299
90	16	1.4	479874	6	US-60-248-505-72
91	16	1.4	582670	6	US-60-248-505-4
92	16	1.4	666480	5	US-09-335-032-12214
93	16	1.4	690793	6	US-60-248-505-227
94	16	1.4	784328	5	US-09-335-032-12217
95	16	1.4	848061	5	US-09-335-032-12219
96	16	1.4	1531974	5	US-09-335-032-12207
97	16	1.4	1531974	5	US-09-335-032-12207
98	15	1.3	125	5	US-09-540-212A-26471
99	15	1.3	128	5	US-09-540-212A-26471
100	15	1.3	151	5	US-09-487-566A-5297

Sequence 9895, Ap
Sequence 31535, A
Sequence 16219, A
Sequence 6168, Ap
Sequence 5608, Ap
Sequence 9337, Ap
Sequence 135, App
Sequence 482, App
Sequence 1755, Ap
Sequence 715, App
Sequence 6254, Ap
Sequence 8336, Ap
Sequence 8338, Ap
Sequence 887, App
Sequence 966, App
Sequence 433, App
Sequence 9, Appl
Sequence 5554, Ap
Sequence 178, App
Sequence 8314, Ap
Sequence 41, Appl
Sequence 359, App
Sequence 179, App
Sequence 238, App
Sequence 357, App
Sequence 2477, Ap
Sequence 8504, Ap
Sequence 8505, Ap
Sequence 3033, Ap
Sequence 3033, Ap
Sequence 3012, Ap
Sequence 1278, Ap
Sequence 628, App
Sequence 9893, Ap
Sequence 16, Appl
Sequence 10125, A
Sequence 1, Appl
Sequence 601, App
Sequence 40, Appl
Sequence 70, Appl
Sequence 8770, Ap
Sequence 3299, Ap
Sequence 370, App
Sequence 157, App
Sequence 501, App
Sequence 351, App
Sequence 85, Appl
Sequence 511, App
Sequence 133, App
Sequence 20, Appl
Sequence 419, App
Sequence 13, Appl
Sequence 14, Appl
Sequence 17, Appl
Sequence 18, Appl
Sequence 2, Appl
Sequence 172, App
Sequence 12206, A
Sequence 215, App
Sequence 28, Appl
Sequence 299, App
Sequence 299, App
Sequence 72, Appl
Sequence 4, Appl
Sequence 12214, A
Sequence 12217, A
Sequence 12219, A
Sequence 12207, A
Sequence 12207, A
Sequence 26471, A
Sequence 7725, Ap
Sequence 5297, Ap

Tue May 15 07:26:57 2001

us-09-151-409-15.oligo.rnpn

Sequence 2784, Ap
Sequence 2790, Ap
Sequence 7, Appli
Sequence 18, Appli
Sequence 9644, Ap
Sequence 553, App
Sequence 312, App
Sequence 392, App
Sequence 587, App
Sequence 393, App
Sequence 358, App
Sequence 8780, Ap
Sequence 3309, Ap
Sequence 409, App
Sequence 8781, Ap
Sequence 3310, Ap
Sequence 6624, Ap
Sequence 822, App
Sequence 2377, Ap
Sequence 551, App
Sequence 531, App
Sequence 7640, Ap
Sequence 9794, Ap
Sequence 332, App

14 1.2 4249 1 PCT-US01-01329-2784
14 1.2 4254 1 PCT-US01-04098A-2790
14 1.2 4286 5 US-09-817-856-7
14 1.2 4305 1 US-09-846-351-18
14 1.2 4370 1 PCT-US01-01339-9644
14 1.2 4399 1 PCT-US01-04098A-553
14 1.2 4446 6 US-60-281-593-312
14 1.2 4694 1 PCT-US01-04098A-392
14 1.2 4709 4 US-08-956-171C-587
14 1.2 4739 1 PCT-US01-04098A-393
14 1.2 4747 1 PCT-US00-35017A-358
14 1.2 4800 1 PCT-US01-01339-8780
14 1.2 4800 1 PCT-US01-01329-3309
14 1.2 4844 1 PCT-US01-04098A-409
14 1.2 4845 1 PCT-US01-01339-8781
14 1.2 4859 1 PCT-US01-01339-6624
14 1.2 4862 1 PCT-US01-04098A-822
14 1.2 4879 1 PCT-US01-04098A-2377
14 1.2 4923 4 US-08-956-171C-551
14 1.2 4945 6 US-60-248-505-531
14 1.2 4964 5 US-09-801-833-7640
14 1.2 5086 1 PCT-US01-01339-9794
14 1.2 5136 4 US-08-956-171C-332
c1000

ALIGNMENTS

RESULT 1
US-09-540-212A-66707
; Sequence 66707, Application US/09540212A
; GENERAL INFORMATION:
; APPLICANT: Sellhame, Jeffrey J.
; APPLICANT: Deleageane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
; FILE REFERENCE: PD-1034 CIP
; CURRENT APPLICATION NUMBER: US/09/540,212A
; CURRENT FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 67551
; SOFTWARE: PERL Program
; SEQ ID NO 66707
; LENGTH: 275
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc:feature
; OTHER INFORMATION: Incyte ID No: hu00684568
US-09-540-212A-66707

Query Match 1.6%; Score 18; DB 5; Length 275;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 152 taactgaagagttgagta 169
Db 169 taactgaagagttgagta 186

RESULT 2
US-08-956-171C-537/c
; Sequence 537, Application US/08956171C
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash

Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171C
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Hoover, Kenley K.
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 610-5790
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 537:
SEQUENCE CHARACTERISTICS:
LENGTH: 784 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 537:
US-08-956-171C-537

Query Match 1.6%; Score 18; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 422 aacttttaacaagtatg 439
Db 673 AACTTCTTACAAAGTATG 656

RESULT 3
PCT-US01-04098A-471
; Sequence 471, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20

```
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 471
; LENGTH: 4813
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (155)..(4813)
PCT-US01-04098A-471

Query Match      1.6%; Score 18; DB 1; Length 4813;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 cagaatgagaagtattct 591
      |||||||
Db 245 cagaatgagaagtattct 262

RESULT 4
PCT-US01-04098A-2439/C
; Sequence 2439, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 2439
; LENGTH: 5623
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US01-04098A-2439

Query Match      1.6%; Score 18; DB 1; Length 5623;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 574 cagaatgagaagtattct 591
      |||||||
Db 5419 CAGAAAGAGAGTATTCT 5402

RESULT 5
US-60-248-505-134
; Sequence 134, Application US/60248505

; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: G1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134
; LENGTH: 283061
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(283061)
; OTHER INFORMATION: n = A,T,C or G
US-60-248-505-134

Query Match      1.6%; Score 18; DB 6; Length 283061;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1113 aaaagctagtaaaattca 1130
      |||||||
Db 280582 aaaagctagtaaaattca 280599

RESULT 6
US-09-540-212A-13409/C
; Sequence 13409, Application US/09540212A
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
; FILE REFERENCE: PD-1034 CIP
; CURRENT APPLICATION NUMBER: US/09/540,212A
; CURRENT FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 67551
; SOFTWARE: PERL Program
; SEQ ID NO 13409
; LENGTH: 132
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu003000062
; NAME/KEY: unsure
; LOCATION: 118
; OTHER INFORMATION: a, t, c, g, or other
US-09-540-212A-13409

Query Match      1.5%; Score 17; DB 5; Length 132;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 519 taagctaaaaaaatta 535
      |||||||
Db 81 TAAGCTAAAAAAATTA 65

RESULT 7
US-09-801-833-7074/C
; Sequence 7074, Application US/09801833
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, M. Alexandra
```

Query Match	1.5%	Score 17	DB 5	Length 1682
Best Local Similarity	100.0%	Pred. No. 36	0	Indels 0; Gaps 0
Matches 17	Conservative	0	Mismatches	0
<p>US-09-801-833-7074</p> <p>Query Match 1.5% Score 17 DB 5 Length 1682</p> <p>Best Local Similarity 100.0% Pred. No. 36</p> <p>Matches 17 Conservative 0 Mismatches 0 Indels 0 Gaps 0</p>				
QY 765	ggttacacaattatata	781		
DB 1315	GGTTACACAATTATATA	1299		
<p>RESULT 8</p> <p>US-09-543-679A-2639/c</p> <p>Sequence 2639, Application US/09543679A</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: NYCE, Jonathan W.</p> <p>TITLE OF INVENTION: LOW ADENOSINE ANTI-SENSE OLIGONUCLEOTIDE, COMPOSITIONS KIT & METHOD FOR TREATMENT OF AIRWAY DISORDERS ASSOCIATED WITH BRONCHOCONSTRICION, LUNG INFLAMMATION, NUMBER OF SEQUENCES: 3111</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: EPIGENESIS PHARMACEUTICALS, INC.</p> <p>STREET: 7 Clarke Drive</p> <p>CITY: Cranbury</p> <p>STATE: NJ</p> <p>COUNTRY: USA</p> <p>ZIP: 08512</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: CD-R</p> <p>COMPUTER: IBM Compatible</p> <p>OPERATING SYSTEM: DOS</p> <p>SOFTWARE: N/A</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/09/543, 679A</p> <p>FILING DATE: 13-Apr-2000</p> <p>CLASSIFICATION: UNKNOWN</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: 60/127, 958</p> <p>FILING DATE: 1998-08-03</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Anzel, Viviana</p> <p>REGISTRATION NUMBER: 30, 930</p> <p>REFERENCE/DOCKET NUMBER: EPI-0067191b</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: 609-409-3035</p> <p>TELEFAX: 413-254-9245</p> <p>TELEX: <Unknown></p> <p>INFORMATION FOR SEQ ID NO: 2639:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 3929 base pairs</p>				

RESULT 10
US-09-543-679A-2640/c
; Sequence 2640, Application US/09543679A
; GENERAL INFORMATION:
; APPLICANT: NYCE, Jonathan W.
; TITLE OF INVENTION: LOW ADENOSINE ANTI-SENSE OLIGONUCLEOTIDE,
; COMPOSITIONS, KIT & METHOD FOR TREATMENT
; OF AIRWAY DISORDERS ASSOCIATED WITH
; BRONCHOCONSTRICTION, LUNG INFLAMMATION,
; NUMBER OF SEQUENCES: 311
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: EPIGENESIS PHARMACEUTICALS, INC.
; STREET: 7 Clarke Drive
; CITY: Cranbury
; STATE: NJ
; COUNTRY: USA
; ZIP: 08512
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: N/A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/543,679A
; FILING DATE: 13-Apr-2000
; CLASSIFICATION: UNKNOWN
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/127,958
; FILING DATE: 1998-08-03
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: EPI-0067191b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-409-3035
; TELEFAX: 413-254-9245
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2640:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2640
US-09-543-679A-2640

Query Match 1.5%; Score 17; DB 5; Length 12447;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 aaactaagaatgaag 293
|||||

Db 10031 AAACTAAGATGAAGG 10015

RESULT 11
US-60-248-505-614/c
; Sequence 614, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: RECEPTORS, AND USES THEREOF
; FILE REFERENCE: c1000518
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 614
; LENGTH: 126883

; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(126883)
; OTHER INFORMATION: n = A,T,C or G
US-60-248-505-614

Query Match 1.5%; Score 17; DB 6; Length 126883;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 ttaagtgttaataataa 125
|||||

Db 26637 TTAAGTTTAAATAATAA 26621

RESULT 12
US-09-335-032-12216
; Sequence 12216, Application US/09335032
; GENERAL INFORMATION:
; APPLICANT: Velculescu, Victor
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: Characterization of the Yeast
; FILE REFERENCE: 01107.78572
; CURRENT APPLICATION NUMBER: US/09/335,032
; CURRENT FILING DATE: 1999-06-16
; PRIOR APPLICATION NUMBER: US 60/035,917
; PRIOR FILING DATE: 1997-01-23
; PRIOR APPLICATION NUMBER: US 09/012,031
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 12219
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12216
; LENGTH: 924430
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-09-335-032-12216

Query Match 1.5%; Score 17; DB 5; Length 924430;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 563 aattacaacaagaat 579
|||||

Db 11585 aattacaacaagaat 115601

RESULT 13
US-09-540-212A-37688/c
; Sequence 37688, Application US/09540212A
; GENERAL INFORMATION:
; APPLICANT: Seilhame, Jeffrey J.
; APPLICANT: Deleage, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullishy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
; FILE REFERENCE: PD-1034 CIP
; CURRENT APPLICATION NUMBER: US/09/540,212A
; CURRENT FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 67551
; SOFTWARE: PERL Program
; SEQ ID NO 37688
; LENGTH: 95
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00858432
US-09-540-212A-37688

Query Match 1.4%; Score 16; DB 5; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 519 taagctaaaaaaatt 534
|||||
Db 68 TAAGCTAAAAAATT 53

RESULT 14
US-09-540-212A-21769/c
; Sequence 21769, Application US/09540212A
; GENERAL INFORMATION:
; APPLICANT: Sellhammer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF AIRWAY AND LUNG SYSTEM TISSUE
; FILE REFERENCE: PD-1034 CIP
; CURRENT APPLICATION NUMBER: US/09/540,212A
; CURRENT FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 67551
; SOFTWARE: PERL Program
; SEQ ID NO 21769
; LENGTH: 120
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00384850
US-09-540-212A-21769

Query Match 1.4%; Score 16; DB 5; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 274 taaaaaactaagaatg 289
|||||
Db 45 TTAATAAACTAAGAATG 30

RESULT 15
US-09-815-343-1168/c
; Sequence 1168, Application US/09815343
; GENERAL INFORMATION:
; APPLICANT: Meagher, Madeleine
; APPLICANT: Xu, Jiangchun
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.504
; CURRENT APPLICATION NUMBER: US/09/815,343
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1168
; LENGTH: 165
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-815-343-1168

Query Match 1.4%; Score 16; DB 5; Length 165;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 600 cggatgatggtatcct 615
|||||
Db 66 CGGTGATGGTAATCCT 51

Search completed: May 13, 2001, 06:26:17
Job time: 3396 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.
OM protein - protein search, using sw model
Run on: May 10, 2001, 08:37:07 ; Search time 44.05 Seconds
(without alignments)
1398.576 Million cell updates/sec
Title: US-09-151-409-16
Perfect score: 1954
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLERASKIQLEARK 383
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1009251 seqs, 160854530 residues
Total number of hits satisfying chosen parameters: 1009251
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Pending Patents AA.Main: *
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US060_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1954	100.0	383	15	US-09-151-409-16
2	312	16.0	443	7	US-08-325-278-6
3	256	13.1	440	17	US-09-302-756-35
4	220.5	11.3	327	11	US-08-761-184-1230
5	220.5	11.3	327	11	US-08-761-184-1230
6	220.5	11.3	327	12	US-08-821-931-1230
7	220.5	11.3	327	12	US-08-821-931-1230
8	220.5	11.3	327	13	US-08-993-002A-8616
9	220.5	11.3	327	13	US-08-993-002A-8617
10	219	11.2	810	16	US-09-248-796-20281, A

11	218.5	11.2	1078	16	US-09-248-796-20284	Sequence 20284, A
12	215.5	11.0	630	16	US-09-248-796-20275	Sequence 20275, A
13	212	10.8	1193	23	US-60-167-217-2860	Sequence 2860, Ap
14	212	10.8	1193	23	US-60-173-464-2336	Sequence 2336, Ap
15	212	10.8	1457	23	US-60-191-637-2817	Sequence 2817, Ap
16	212	10.8	1457	23	US-60-191-681-2256	Sequence 2256, Ap
17	207.5	10.6	980	1	PCT-US99-26796-152	Sequence 152, Appl
18	206.5	10.6	2349	1	PCT-US99-17738-68	Sequence 68, Appl
19	203	10.4	42	1	PCT-US94-12268-31	Sequence 31, Appl
20	203	10.4	42	5	US-08-143-412A-31	Sequence 31, Appl
21	203	10.4	42	8	US-08-479-690-31	Sequence 16333, A
22	201	10.3	905	16	US-09-248-796-16333	Sequence 1, Appli
23	200	10.2	1939	17	US-09-310-187-1	Sequence 1, Appli
24	200	10.2	1939	17	US-09-310-187A-1	Sequence 1, Appli
25	198.5	10.2	2954	15	US-09-150-867-1	Sequence 1, Appli
26	198.5	10.2	2954	19	US-09-536-933-1	Sequence 1, Appli
27	196	10.0	1948	23	US-60-259-128-4790	Sequence 4790, Ap
28	196	10.0	1979	1	PCT-US99-26796-28	Sequence 28, Appl
29	195.5	10.0	1788	18	US-09-488-725A-5398	Sequence 5398, Ap
30	195	10.0	1381	23	US-60-191-637-552	Sequence 552, Appl
31	195	10.0	1381	23	US-60-191-681-435	Sequence 435, Appl
32	194.5	10.0	2688	18	US-09-488-725A-5814	Sequence 5814, Ap
33	193.5	9.9	561	13	US-08-967-909-2	Sequence 2, Appli
34	193	9.9	722	16	US-09-248-796-20613	Sequence 20613, A
35	191.5	9.8	2663	18	US-09-488-725A-2242	Sequence 2242, Ap
36	190.5	9.7	753	18	US-09-458-113-2	Sequence 2, Appli
37	190.5	9.7	1690	23	US-60-167-217-10243	Sequence 10243, A
38	190.5	9.7	1690	23	US-60-167-217-10339	Sequence 10339, A
39	190.5	9.7	1690	23	US-60-173-464-8250	Sequence 8250, Ap
40	190.5	9.7	1690	23	US-60-173-464-8335	Sequence 8335, Ap
41	190.5	9.7	1690	23	US-60-191-637-10256	Sequence 10256, A
42	190.5	9.7	1690	23	US-60-191-637-10343	Sequence 10343, A
43	190.5	9.7	1690	23	US-60-191-681-7997	Sequence 7997, Ap
44	190.5	9.7	1690	23	US-60-191-681-8069	Sequence 8069, Ap
45	190.5	9.7	1780	18	US-09-488-725A-1826	Sequence 1826, Ap

ALIGNMENTS

RESULT 1
US-09-151-409-16
; Sequence 16, Application US/09151409A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: GROUP A STREPTOCOCCAL VACCINES
; FILE REFERENCE: 481112.410
; CURRENT APPLICATION NUMBER: US/09/151,409A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: 60/058,635
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Hexavalent M
; OTHER INFORMATION: fusion gene sequence constructed from
; OTHER INFORMATION: streptococcal type 24, 5, 6, 19, 1 and 3 M protein
; OTHER INFORMATION: DNAs
US-09-151-409-16

Query Match 100.0%; Score 1954; DB 15; Length 383;
Best Local Similarity 100.0%; Pred. No. 6e-127;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACWATRSQDTLEKVOERADKFEIENNTLKNSDLSPFNKALKHNDLFEELSNAKE 60
|||||
Db 1 ACWATRSQDTLEKVOERADKFEIENNTLKNSDLSPFNKALKHNDLFEELSNAKE 60

```
QY 61 KLRGSAVTRGTINDPQAKALDYELNHDLTCKNEGLKTENEGTKTENEGL 120
DB 61 KLRGSAVTRGTINDPQAKALDYELNHDLTCKNEGLKTENEGTKTENEGL 120
QY 121 KTEVDVFPRTGTVENPKARELLNKYDVNSMLQANNKLPWRVRYTRHTPEDKLLKIID 180
DB 121 KTEVDVFPRTGTVENPKARELLNKYDVNSMLQANNKLPWRVRYTRHTPEDKLLKIID 180
QY 181 DLDAKEHELOQOONKLSLQNGDGNPREVIEDLAANNPAIONIRLHENKDKLKALENAM 240
DB 181 DLDAKEHELOQOONKLSLQNGDGNPREVIEDLAANNPAIONIRLHENKDKLKALENAM 240
QY 241 VAGROFKRAGTLLDQVTLTKHNSNTQYNAQAAGRLDLROKAEYKLGKLNDAERLLQEL 300
DB 241 VAGROFKRAGTLLDQVTLTKHNSNTQYNAQAAGRLDLROKAEYKLGKLNDAERLLQEL 300
QY 301 NIDVATRSQDTLEKVQBRADKFEIENNTLKLKNSDLSFNKALKOHNDLDELTELSNAKE 360
DB 301 NIDVATRSQDTLEKVQBRADKFEIENNTLKLKNSDLSFNKALKOHNDLDELTELSNAKE 360
QY 361 KLRKNDKLSSEKASKIOLEARK 383
DB 361 KLRKNDKLSSEKASKIOLEARK 383

RESULT 2
US-08-325-278-6
; Sequence 6, Application US/08325278
; GENERAL INFORMATION:
; APPLICANT: Bjvick, Lars
; APPLICANT: SJVICK, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/325,278
; FILING DATE: 26-OCT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 450023.401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-325-278-6

Query Match 16.0%; Score 312; DB 7; Length 443;
Best Local Similarity 43.1%; Pred. No. 1.5e-13;
Matches 93; Conservative 18; Mismatches 65; Indels 40; Gaps 7;

QY 200 NCDGNPREVIEDLAANNPAIONIRLHENKDKLKALENAMVAGRDFKRA-----GTL 252
DB 1 NCDGNPREVIEDLAANNPAIONIRLHENKDKLKALENAMVAGRDFKRAEELEKAKOAL 60

QY 253 LDQVTLTKHNSNYQY-----NAQAAGRLDLROKAEYKLGKLNDAER----- 295
DB 61 EDORCKLETKLQLODYDLAKESTSWDRQLEKELEKEKEALELAIDOASRDYHRATAL 120
QY 296 -----LQELNIDVATR--SQTDTLEKVQBRADK--FEIENNTL-----KLKNSDLSFN 341
DB 121 EKELEEKKALELAIDOASQDYNRANVLEKELETITREQEINRNLLGNNAKLELDQLSSEK 180
QY 342 KALKDHNDLTEE--LSNAKEKLRKNDKLSSEKASK 375
DB 181 EQLTIERAKLEEEKQISDASRSQSLRRDLSDASREAKK 216

RESULT 3
US-09-302-756-35
; Sequence 35, Application US/09302756
; GENERAL INFORMATION:
; APPLICANT: FISCHETTI, Vincent A.
; APPLICANT: POZZI, Gianni
; APPLICANT: SCHNEEWIND, Olaf
; TITLE OF INVENTION: DELIVERY AND EXPRESSION OF A HYBRID SURFACE PROTEIN ON
; TITLE OF INVENTION: THE SURFACE OF GRAM POSITIVE BACTERIA
; FILE REFERENCE: 016921-076
; CURRENT APPLICATION NUMBER: US/09/302,756
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: US 07/522,440
; EARLIER FILING DATE: 1990-05-11
; EARLIER APPLICATION NUMBER: US 07/742,199
; EARLIER FILING DATE: 1991-08-05
; EARLIER APPLICATION NUMBER: US 07/814,823
; EARLIER FILING DATE: 1991-12-23
; EARLIER APPLICATION NUMBER: US 07/851,082
; EARLIER FILING DATE: 1992-03-13
; EARLIER APPLICATION NUMBER: PCT/US93/02355
; EARLIER FILING DATE: 1993-03-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 35
; LENGTH: 440
; TYPE: PRT
; ORGANISM: S. pyogenes
; US-09-302-756-35

Query Match 13.1%; Score 256; DB 17; Length 440;
Best Local Similarity 30.8%; Pred. No. 1.1e-09;
Matches 100; Conservative 34; Mismatches 111; Indels 80; Gaps 12;

QY 126 RVFPRGTVENPKARELLNKYDVNSMLQANNKLPWRVRYTRHTPEDKLLKIIDDLAK 185
DB 1 RVFPRGTVENPKARELLNKYDVNSMLQANNKLT-----TENNLTQNKNLTT 52
QY 186 EHELOQOONKLSLQNGDGNPREVIEDLAANNPAI--QNIRLHENKDKLKALENAMVAG 243
DB 53 NKNLTQNKNLTTEN--KNLTQNKNLTTENKELKAEGRNLTENKGLTKLSEAESEAA 110
QY 244 ---RDFKRA--GTLDDQVTLTKHNSNYQYNAQAAGRL-----DLROKAEYKLGKLN 290
DB 111 NKERENKEAIGTLTKTLDLDETVDKDKIAKEQESKETIGTLKLTDETVKDKIAKEQESKETI 170
QY 291 DWAERLLQELNIDVATRSTQD-----TLKV-----QERADK 322
DB 171 GTLTKTLDLDETVDKDKIAKEQESKETIGTLKLTDETVKDKIAKEQSKQDICALKQELAKK 230
QY 323 FE---TENNTLKLKNSDLSFNKALKDHNDL-----LITEELLSNAKE-----LRKN- 365
DB 231 DEGNKVSASRKGRLRRDLSDASREAKKQVEKDLANLTAEKDKVKEEKQISDASRQGLRRDL 290
QY 366 -----DKSLSEKASKIOELE 380
DB 291 DASREAKKQVEKALEEANSKLALE 315
```

```

RESULT      4
US-08-761-184-1230
; Sequence 1230, Application US/08761184
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH ET AL
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND
; TITLE OF INVENTION: THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 1810
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/761,184
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,032
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/561,469
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,405
; FILING DATE: 01-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/660,742
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/736,791
; FILING DATE: 25-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/739,150
; FILING DATE: 28-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-001CP9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 1230:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...327
US-08-761-184-1230

Query Match          11.3%; Score 220.5; DB 11; Length 327;
Best Local Similarity 25.7%; Pred. No. 1.9e-07;
Matches 97; Conservative 64; Mismatches 133; Indels 83; Gaps 15;

QY 14 EKVGQRADKEFEIENNNTLKKNSDLSFNKKDKHNDEL---TEELSNAEKILRGSAVTG 70
::: | : | | | : | : | | | | | : | | |
Db 16 EELEARISELENTELLRREREYLAAETSEKDAOQLRKNDKLFITDKDL----- 67
::: | : | | | : | : | | | | | : | | |
QY 71 TINDPQRAEKALDYELNHDLTKNEGKLTNEGLKTNEGIKTVEORVFPR 130

```

:	INFORMATION FOR SEQ ID NO:	1445;
:	SEQUENCE CHARACTERISTICS:	
:	LENGTH: 327 amino acids	
:	TYPE: amino acid	
:	TOPOLOGY: linear	
:	MOLECULE TYPE: protein	
:	HYPOTHETICAL: YES	
:	ORIGINAL SOURCE:	
:	ORGANISM: Helicobacter pylori	
:	FEATURE:	
:	NAME/KEY: misc_feature	
:	LOCATION: 1...327	
:	US-08-761-184-1445	
:	Query Match	11.3%; Score 220.5; DB 11; Length 327;
:	Best Local Similarity	25.7%; Pred. No. 1.9e-07;
:	Matches	97; Conservative 64; Mismatches 133; Indels 83; Gaps 15;
QY	14 EKQVERADKFEIENNTLKLNSDLSPNKKALDHDEL---TEELSNAAKEKLRGSAVTRG	70
DB	16 EELEARISELEDENTELLREREYLAAETSELKDANDQLRKNDKLFITDKRL-----	67
QY	71 TINDPQAKALKYLENHDLTKNEGLKTENEGLKTENEGLTDEVDVFPR	130
DB	68 -----TKENTELF-AENESLSVKISGLEHSDQLWQNKNLTKEKAELKTE-----	112
QY	131 GTVENPDKARELNKYDVENSIMLAQNNDKLPWRVRYTHPTPEKLIKIIDDDAKEHELQ	190
DB	113 -----KDILAK---ENTRLAARDRLTEEKRELT--TEKERLKRENTTELTHKITELT	159
QY	191 QONEKLSLONGDNPREVEDLAANNPATONIRLHENKDKARLENAMAVAGDFKRAG	250
DB	160 KENKALTEN-----DKNHQVTALTNERDSLEQE--RARLODA-----HG	198
QY	251 TLDDQTOLVTKHNSNYQQYNAQAGRIDLQRKAAYLYKGLNDWAERLLQELNIDIVATRSQT	310
DB	199 FLEXRCNL-EKEN--ORLTDKIKLESQAQS--LENTNNQLRALSENSVOLAQAEK	252
QY	311 DTLEKVOERADKFEIENNTLKLKN-----SLSFNKALKDHDELTEELSNAKEK--	361
DB	253 IAIE-----KSELEREIALKSLEGMEAKSLDLHNRRLASANEDLRQNRKLEEENI	305
QY	362 -LRKNDKLSLEKASKIQ	377
DB	306 ALKERVDCNLQSLKIQ	322
:	RESULT	6
:	US-08-821-931-1230	
:	Sequence 1230, Application US/08821931	
:	GENERAL INFORMATION:	
:	APPLICANT: DOUGLAS SMITH ET AL	
:	TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES	
:	TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND	
:	TITLE OF INVENTION: THERAPEUTIC USES THEREOF	
:	NUMBER OF SEQUENCES: 1810	
:	CORRESPONDENCE ADDRESS:	
:	ADDRESSEE: LAHIVE & COCKFIELD	
:	STREET: 60 State Street, Suite 510	
:	CITY: Boston	
:	STATE: Massachusetts	
:	COUNTRY: USA	
:	ZIP: 02109-1875	
:	COMPUTER READABLE FORM:	
:	MEDIUM TYPE: CD-ROM ISO9660	
:	COMPUTER:	
:	OPERATING SYSTEM:	
:	SOFTWARE:	
:	CURRENT APPLICATION DATA:	
:	APPLICATION NUMBER: US/08/821,931	
:	FILING DATE:	
:	PRIOR APPLICATION DATA:	

RESULT 7
 US-08-821-931-1445
 ; Sequence 1445, Application US/08821931
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH ET AL
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI AND
 ; TITLE OF INVENTION: THERAPEUTIC USES THEREOF
 ; NUMBER OF SEQUENCES: 1810
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, Suite 510
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD/ROM ISO9660
 ; COMPUTER:
 ; OPERATING SYSTEM:
 ; SOFTWARE:
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/821,931
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/487,032
 ; FILING DATE: 07-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/561,469
 ; FILING DATE: 17-NOV-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/630,405
 ; FILING DATE: 01-APR-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/660,742
 ; FILING DATE: 06-JUN-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/736,791
 ; FILING DATE: 25-OCT-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/739,150
 ; FILING DATE: 28-OCT-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/761,184
 ; FILING DATE: 06-DEC-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mandragouras, Amy E.
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: GTN-001CP9CN
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)227-5941
 ; INFORMATION FOR SEQ ID NO: 1445:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 327 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Helicobacter pylori
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1...327
 US-08-821-931-1445

Query Match 11.3%; Score 220.5; DB 12; Length 327;
 Best Local Similarity 25.7%; Pred. No. 1.9e-07;
 Matches 97; Conservative 64; Mismatches 133; Indels 83; Gaps 15;

QY 14 EKVOERADKFEIENNTLKLKNSDLSPNNKALKDHNDL---TELSNAKEKLRGSVTRG 70
 Db 16 EELEARISEDENTELLRREREYLAETSELKDANDQLRQKNDKLFITDKL----- 67
 QY 71 TINDPQRAKEALDKYELENHDLTKNEGLKTENEGLKTENEGKLTKEVDRVFPFR 130
 Db 68 -----TKENTELF-AENESLSVKISGLEHSDQWNNKLTKEAEKTE----- 112
 QY 131 GTVENPKARELLNKYDVENSMLQANNDKLPWRVRYTRHTPEDKLLKIIDDLDAKEHELQ 190
 Db 113 -----KDILAK---ENTRLAARDRLTEKRELT-TEKERLKRENTLTHITELT 159
 QY 191 QQNEKLSLQNGDGNPREVIEDLAANPAONIRLHKNKDLKARLENAMEVAGRDFKRAK 250
 Db 160 KENKALTEN-----DKLNHQVTALTNERDSLEQF--RARIQDA-----HG 198
 QY 251 TLDDQVTQLYTKHNSYQYNAQAGRLDLRQAEYLLKGLNDWAERLLQELNIDVATRSQT 310
 Db 199 FLEKRCNLT-EKEN---QRLTDKLLQLESQAQS--LENTNQLRQALENSVQLAQAEK 252
 QY 311 DTLEKVOERADKFEIENNTLKLKN-----SDLSFNKALKDHNDLTELSELSNAKEK-- 361
 Db 253 IAE-----KSELERETARUKSLEGMKASDLDLHNRRLASANEDLKRONKLEENI 305
 QY 362 -LRKNDKSLSEKASKIQ 377
 Db 306 ALKERVDTGLNEQLSKLQ 322
 RESULT 8
 US-08-993-002A-8616
 ; Sequence 8616, Application US/08993002A
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGLAS SMITH et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
 ; TITLE OF INVENTION: RELATING TO HELICOBACTER PYLORI FOR
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 10031
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 28 State Street
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD/ROM ISO9660
 ; COMPUTER: IBM PC Compatible
 ; OPERATING SYSTEM: Windows NT 4.0
 ; SOFTWARE: UNIX
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/993,002A
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mandragouras, Amy E.
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: GTN-018
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)742-4214
 ; INFORMATION FOR SEQ ID NO: 8616:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 327 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Helicobacter pylori
 ; FEATURE:

```

; NAME/KEY: misc.feature
; LOCATION: 1...327
; US-08-993-002A-8616

Query Match      11.3%; Score 220.5; DB 13; Length 327;
Best Local Similarity 25.7%; Pred. No. 1.9e-07;
Matches 97; Conservative 64; Mismatches 133; Indels 83; Gaps 15;

QY 14 EKQVGRADKFEIENNTLKLNSDLSPFNKALKDHNDL---TEELSNKAEKLRGSVTRG 70
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 16 EELEARISELEDENTELLREREYLAETSELKDANDOLROKNDKLFITDKL----- 67
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 71 TINDPQRAKEALDKYLENHDLTKNEGLKTENEGTKTENEGTKTEVDRVFR 130
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 68 -----TKENTELF-AENESLSVKISLEHSNQLQNNKLTKEKAELKTE----- 112
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 131 GTVENPDKARELLNKYDVENSMLQANDKLPWRVRYTRHTPEDKLLIIDLDKAEHELQ 190
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 113 -----KDILAK---ENTRLAARDLTTEKRELT-TEKERLKRNTTELTHKITELT 159
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 191 QONEKLSLONQGNPREVEDLAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKFRAG 250
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 160 KENKALTEN-----DKLNHQVTALTNERDSLEGE---RARLQDA-----HG 198
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 251 TLDDQVTLQYTKHNSYQYNAQAGRLDLROKAEYKGLNDWAERLLQELNIDVATRSQT 310
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 199 FLEKRCNTL-EKEN---QRLTDKQLKQESAKS---LENTNNQLQALENSNOVLAQAKEK 252
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 311 DTLEKVOERADKFEIENNTLKLKN-----SDLSFNKALKDHNDLTELKNAKEK-- 361
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 253 IAIE-----KSELERETARLSLEGMEAKSDLDLHNRRLASANEDLKQRNKRLEENI 305
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

QY 362 -LRNKDKSLSEKASKIQ 377
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 306 ALKERVGDGLNEQLSKLQ 322
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

RESULT 10
US-09-248-796-20281
; Sequence 20281, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20281
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796-20281

Query Match      11.2%; Score 219; DB 16; Length 810;
Best Local Similarity 23.1%; Pred. No. 9.3e-07;
Matches 116; Conservative 80; Mismatches 146; Indels 160; Gaps 23;

QY 6 TRSQDT--LEKVOERADKFEIENNTLKLNSDLSPFNKALKDHNDLTELK--- 54
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 186 TLPETDTNEVEIKSIEDNLTENNRSLATEENDNDNEEDDDDDITKESRGSTVYHQ 245
   |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|

```



```
QY 55 ---ISNA-----KEKLRGSAVTRGTINDPQRAKALDKYLENHDLDKTKNEGLKTENG 105
Db 246 NOTLDNTQISRIREDLKDYIT-----PERATEVAQSPHPHQHRIELIEDKTEDFG 299
QY 106 -----LKTENEGKTENE-----GLKTEYDVRFPFGTGVENPDKARELLNK 145
Db 300 KFNDTTSPKTRNQ---DENEFDETANLTLLRQLENEPTINFLMSPNSKPVFSKQVYNK 356
QY 146 YDVE-----NSMLQANNDKLPWRVYTRHTPDKLAKIIDDL-----DAKEHELOOQ 192
Db 357 IQEHESKVDTLQLQNSR-----DDKIKOLYDELAKVNEQLISQKQORALN 404
QY 193 NEKLSLQNGDG-----NPREV-----IEDLAANN-----PAIQNIRLRHKN 229
Db 405 DEKSLANSENLLTIQHHNERELASLTKSFRLKESLNNLKRKLNEQKVSIDETRHEND 464
QY 230 DLKARLENAMAVAGDRFKR-----AGTL-LDQV-----TQLYTKHNSYQOYNQA 274
Db 465 TLKSKI-NEQSLERDLKAEIVETSNTSTESTLKLDQVKEKEBELWIKN---QELNEKL 519
QY 275 GRDLRQKAEYKGLNDWAERLLOELINIDVATRSQTDLEKVOERADKFEIENNTLKI-- 332
Db 520 QSFESNSSKYDK-LKDEYEDLLE-----RNQTLVDDNKRLETDKFELEDRTKKLEA 570
QY 333 ---KNSDLSFNN-----KALKDNDDELTEELS-----NAKE 360
Db 571 VNQKLADDNTNNENLIKEYKAVATRIELESERDSTIEISKHKQIVHDIETQVENLKQ 630
QY 361 KLRKNDKSLSEKASKIOLEAR 382
Db 631 OYNDKEKTNKELELRIELESK 652

RESULT 11
US-09-248-796-20284
; Sequence 20284, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20284
; LENGTH: 1078
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-20284

Query Match 11.2%; Score 218.5; DB 16; Length 1078;
Best Local Similarity 24.1%; Pred. No. 1.5e-06;
Matches 105; Conservative 81; Mismatches 172; Indels 77; Gaps 19;

QY 1 ACWATRSQTDTL-----EKVQERADKFEIENNTLKNKSDLSFNKA---LKHNDDEL 52
Db 436 SAMKASKFEFETASQYKQELQALAKGNTSESTLKQLEKLDSTEQAKKLEDEINMT 495
QY 53 BELSNAKEKLRGSAVTRGTINDPQRAKALD-KYLEN-----HDLTKNEGLKTE 102
Db 496 RDLFHLK-KLKSAETQ-----IKQREPFKNLTYELENTKDYELQINNKNSENFQK 550
QY 103 NEGLKTENEGKTENGKTEYDVRFPFGTGVENPDKARELLNKYDVE-NSMLQANNDKLP 161
Db 551 INELSKKIESLTEDNKSNAQLEEK-LRTEENNEHMDKLRASAVAYNDLKAKSESE 609
QY 162 WRVYTRHTPDKLAKIIDDLDAKEHELOOQNEKLS-----LONGDGNPREVIEDL 212
Db 610 ETYK-----AKEEFTLTSKIDNLEKEQKQKNELEGQLQNTIDSTNEKFELEDEL 664
QY 213 -----AANNPAIQNIRLRHENDKLRLENAMAVAGDRFKRAGTILLDQVTOYLT 261
```

```
Db 665 KSIKSNKEISSONSELIQ--KLEKTEKLOAKDEIDKKAETKSNIDNLSNLISSLOS 722
QY 262 K-----HNSNTQOYNQAAGRLDLROKAEYKGLNDWAERLLOEL-----NIDVAT 306
Db 723 KLKAEESHSSSTKDEHSSLSSENK-KLKEEYENTKTSMTAKLSAKIEEHKATDIEYTK 781
QY 307 RSQTDLEKVOERADKFEIENNTLKNKSDLSFNKALKDNDDELTEELSN-AKEKLRKN 365
Db 782 KHITDLOEBAKQKQSFESERNDIK---SNLDEANKELSDNR-----EKLNSLEKEKTELN 834
QY 366 DKSLSSEKASKIOELE 380
Db 835 NK-LKTQEEKISDLE 848

RESULT 12
US-09-248-796-20275
; Sequence 20275, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 20275
; LENGTH: 630
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796-20275

Query Match 11.0%; Score 215.5; DB 16; Length 630;
Best Local Similarity 23.1%; Pred. No. 1.1e-06;
Matches 104; Conservative 92; Mismatches 152; Indels 103; Gaps 19;

QY 7 RSQTDLEKVOERADKFEIENNTLKNKSDLSFNKALKDNDDEL-TEELSNKAEKLRGS 65
Db 110 KHNTESATSEKNNQIKELSETIKSLTKTSGDALQSQKQYKTLTKNSDTELK-- 167
QY 66 AVTRGTINDPQRAKALDKYLENHDLDKTKNEGLK--TENE-GLKTENEGKTENEGKLT 122
Db 168 -----LEKQLELEKVKL---DLQTADEKLGITEREIALKSELETYK--NSGLST 213
QY 123 EVDVRFPGTGVENPDKARELL-----NKYDVNSMLQANNDKLPWRVYTRHTPDKLKK 177
Db 214 TSELAALTQTVKSLKEKEELOFLSGNKKSELDYIQKHSD-----ISEKKA 261
QY 178 IIDDLDAKEHELOOQNEKLS-LONGDGNPREVIEDLAANNPAIQNIRLRH 230
Db 262 LTDELKREKTQFDDSKKLTLENDLTSTKKELETEKTQTSKFKNLEERKQKEIVKLANKE 321
QY 231 LK-----ARLENAMAVAGDRFKRAGTILLDQVTOL---YTKHNSYQO-- 269
Db 322 LELLKNDNSGAKKELLEKVKSEISEIILSKELEDKSKVMQHDDELKEQTKKNOELQKV 381
QY 270 ---YNAQAGRLDLROK-----AEYKGLNDWAERLLOELNIDVATRSQ 310
Db 382 TKDYSTTKLLELOKELDAALSFKDKFETASAKLTQSTSD-LEAANKKLNILISEK 440
QY 311 D-TLEKVOERADKFE-----IENNTLKNKSDLSFNKALKD---HNDELTEELSNK 362
Db 441 EQLEKLTQKHALEKLIIDKNADSSKTSNDSLEKANKKVLDETQLEKLTDELDTNKL 500
QY 363 RKN-----KSLSEKASKIOELEARK 383
Db 501 QXHEKLNNEFTNLSKHEEIKTKDKIENEK 531

RESULT 13
US-60-167-217-2860
; Sequence 2860, Application US/60167217
```

```

; LOCATION: (1)....(1193)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-60-173-464-2336

Query Match 10.8%; Score 212; DB 23; Length 1193;
Best Local Similarity 26.2%; Pred. No. 5e-06;
Matches 110; Conservative 66; Mismatches 176; Indels 68; Gaps 16;

Qy 7 RSQDTLEKVOERADKFEIENNTLK--LKNSDLSFNKKALKDHNDELTEELSNAEKL-- 62
Db 269 RSSAD-WESTKQRIARLELENERLKHDLERSQTTFGRTM-----TTSQELDRAQERADK 322
Qy 63 -----RGSVTRGTINDPORAK-----EALDKYELENHDLKTKNEGLKTENEGLKTE 109
Db 323 ASAEELRRQTQAELRVTQSDAEREEAAALQEKLEKSGEYVRLKAKLENAQGEQESLRQE 382
Qy 110 NEGLKTENEGLKTEVDRAVEP--RGTVENPDKARELLNKYDVNSMLQANDKLPMRVRYT 167
Db 383 LEKAQSGVSRHADRDRAFSEVEKIKEMEERTQATLGKSQLQHEKLQNSLDKAQNEVDHL 442
Qy 168 RHTPEDKL-----KKIIDDLDAKHEHL-----QQQNEKLSLQONGDG 203
Db 443 ----QDKLDKACTENRRLVLEKEKLTVDYNLSQSLDKALGQAARMQKERETLSLOT--D 496
Qy 204 NPREVIEDLAANNFAIGNIR--LRHENKDLKARLENAMEV---AGRDFKRAGLTLLDQVTO 258
Db 497 RIRKELEKTQVLGRIGKQERDOFSDELETKERSESAQTLMLKAARDREAMQTDLEVLKE 556
Qy 259 LYTHNSNYOQYNAQAQRLDLRQKAAYLGLNDWAERLLQELNLIDIVATRSQDTLEKVOE 318
Db 557 RYEK--SHAIQOKLQMERDDAVTEVILKEKLDKALYASOKL-IDEKDTSNKE-FEKLME 612
Qy 319 RADKFEIENNTLKNSDLSFNKKALKDHNDELTEELSNAEKLKRNKDKSLSEKASKIQE 378
Db 613 KYDRAQNEIYRLQSCDTAEADRARLEVAERSGLAASKAREDLRK----LQDESTRLQE 668

RESULT 15
US-60-191-637-2817
; Sequence 2817, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO00392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2817
; LENGTH: 1457
; TYPE: PRT
; ORGANISM: DROSOPHILA
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)....(1457)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-60-191-637-2817

Query Match 10.8%; Score 212; DB 23; Length 1457;
Best Local Similarity 26.2%; Pred. No. 6.7e-06;
Matches 110; Conservative 66; Mismatches 176; Indels 68; Gaps 16;

Qy 7 RSQDTLEKVOERADKFEIENNTLK--LKNSDLSFNKKALKDHNDELTEELSNAEKL-- 62
Db 269 RSSAD-WESTKQRIARLELENERLKHDLERSQTTFGRTM-----TTSQELDRAQERADK 322
Qy 63 -----RGSVTRGTINDPORAK-----EALDKYELENHDLKTKNEGLKTENEGLKTE 109
Db 323 ASAEELRRQTQAELRVTQSDAEREEAAALQEKLEKSGEYVRLKAKLENAQGEQESLRQE 382
Qy 110 NEGLKTENEGLKTEVDRAVEP--RGTVENPDKARELLNKYDVNSMLQANDKLPMRVRYT 167
Db 383 LEKAQSGVSRHADRDRAFSEVEKIKEMEERTQATLGKSQLQHEKLQNSLDKAQNEVDHL 442
Qy 168 RHTPEDKL-----KKIIDDLDAKHEHL-----QQQNEKLSLQONGDG 203
Db 443 ----QDKLDKACTENRRLVLEKEKLTVDYNLSQSLDKALGQAARMQKERETLSLOT--D 496
Qy 204 NPREVIEDLAANNFAIGNIR--LRHENKDLKARLENAMEV---AGRDFKRAGLTLLDQVTO 258
Db 497 RIRKELEKTQVLGRIGKQERDOFSDELETKERSESAQTLMLKAARDREAMQTDLEVLKE 556
Qy 259 LYTHNSNYOQYNAQAQRLDLRQKAAYLGLNDWAERLLQELNLIDIVATRSQDTLEKVOE 318
Db 557 RYEK--SHAIQOKLQMERDDAVTEVILKEKLDKALYASOKL-IDEKDTSNKE-FEKLME 612
Qy 319 RADKFEIENNTLKNSDLSFNKKALKDHNDELTEELSNAEKLKRNKDKSLSEKASKIQE 378
Db 613 KYDRAQNEIYRLQSCDTAEADRARLEVAERSGLAASKAREDLRK----LQDESTRLQE 668

```

Search completed: May 10, 2001, 08:39:39
Job time: 152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:08 ; Search time 31.03 Seconds
(without alignments)
84.257 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 1954
Sequence: 1 ACMVATRSQDTLEKVOERA.....KNDKSLSEKAIQLEARK 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 39611 seqs, 6826361 residues

Total number of hits satisfying chosen parameters: 39611

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	256	13.1	440	4	US-08-302-756C-35
2	207.5	10.6	931	1	PCT-US01-04098A-3150
3	207.5	10.6	990	1	PCT-US01-04098A-1182
4	184	9.4	1879	1	PCT-US00-35017A-1265
5	177.5	9.1	1047	1	PCT-US01-04098A-1169
6	177.5	9.1	1047	1	PCT-US01-04098A-3137
7	176	9.0	2816	1	PCT-US01-10484-145
8	175	9.0	1960	1	PCT-US01-04098A-1516
9	174	8.9	1963	1	PCT-US01-04098A-3484
10	173	8.9	1203	1	PCT-US01-04098A-1926
11	170.5	8.7	484	1	PCT-US01-04098A-1647
12	170.5	8.7	533	1	PCT-US01-04098A-3615
13	165	8.4	902	6	US-60-248-505-973
14	163.5	8.4	992	6	US-60-248-505-1112
15	155	7.9	456	1	PCT-US01-10484-76
16	152	7.8	502	1	PCT-US01-03782A-1066
17	151.5	7.8	954	1	PCT-US01-03782A-108
18	151.5	7.8	1193	1	PCT-US00-35017A-1117
19	151	7.7	1002	5	US-09-834-287A-475
20	151	7.7	1002	5	US-09-834-287A-475
21	151	7.7	1095	5	US-09-834-759-493
22	146.5	7.5	432	5	US-09-825-301-18
23	146.5	7.5	432	5	US-09-604-287A-181
24	145.5	7.4	1275	1	PCT-US01-04098A-1583
25	145.5	7.4	1301	1	PCT-US01-04098A-3551
26	145.5	7.4	963	1	PCT-US01-04098A-1542
27	143.5	7.3			

28	143.5	7.3	979	1	PCT-US01-04098A-3510
29	140.5	7.2	1160	1	PCT-US01-04098A-3188
30	140	7.2	612	1	PCT-US01-04098A-1800
31	140	7.2	690	1	PCT-US01-04098A-3768
32	139.5	7.1	1179	1	PCT-US01-04098A-1220
33	139	7.1	2515	1	PCT-US01-04098A-3914
34	139	7.1	2515	1	PCT-US01-04098A-3915
35	138.5	7.1	445	5	US-09-825-301-29
36	138.5	7.1	445	5	US-09-604-287A-473
37	138.5	7.1	445	5	US-09-834-759-473
38	135	6.9	2128	6	US-60-248-505-927
39	133.5	6.8	855	1	PCT-US01-04098A-1128
40	133.5	6.8	873	1	PCT-US01-04098A-3096
41	133	6.8	977	1	PCT-US01-04098A-1201
42	133	6.8	1003	1	PCT-US01-04098A-3169
43	133	6.8	1053	5	US-09-724-519-2
44	132.5	6.8	779	1	PCT-US00-35017A-856
45	131.5	6.7	710	1	PCT-US01-04098A-1159

ALIGNMENTS

RESULT 1
US-08-302-756C-35
; Sequence 35, Application US/08302756C
; GENERAL INFORMATION:
; APPLICANT: FISCHETTI, Vincent A.
; APPLICANT: POZZI, Gianni
; APPLICANT: SCHNEEWIND, Olaf
; TITLE OF INVENTION: DELIVERY AND EXPRESSION OF A HYBRID SURFACE PROTEIN ON
; FILE REFERENCE: THE SURFACE OF GRAM POSITIVE BACTERIA
; CURRENT APPLICATION NUMBER: 016921-076
; CURRENT FILING DATE: 1995-03-07
; PRIOR APPLICATION NUMBER: US/08/302,756C
; PRIOR FILING DATE: 1990-05-11
; PRIOR APPLICATION NUMBER: US 07/742,199
; PRIOR FILING DATE: 1991-08-05
; PRIOR APPLICATION NUMBER: US 07/814,823
; PRIOR FILING DATE: 1991-12-23
; PRIOR APPLICATION NUMBER: US 07/851,082
; PRIOR FILING DATE: 1992-03-13
; PRIOR APPLICATION NUMBER: PCT/US93/02355
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 35
; LENGTH: 440
; TYPE: PRT
; ORGANISM: S. pyogenes
US-08-302-756C-35

Query Match	13.1%	Score 256:	DB 4:	Length 440:
Best Local Similarity	30.8%	Pred. No.	1.8e-09:	
Matches 100;	Conservative	34;	Mismatches 111;	Indels
80;	Gaps	12;		
QY	126	RVFPGTGVNPDKARELLNKYDVNSMLQANNDKLPWRVRYTRHTPEDKLIKTIIDDLDAK	185	
DB	1	RVFPGTGVNPDKARELLNKYDVNSMLQANNDKLT-----TENNLTQDNKNLTTE	52	
QY	186	EHELOQQNEKLSLQNGDNPREVIEDLAANNPAI--QNTLRHENKDLKARLENAMVAG	243	
DB	53	KNKLTQDNKNLTTE--KNLTQDNKNLTTEKNELKAGNRLTTEKGLTKLSEAEDEEA	110	
QY	244	---RDFKRA--GTLDDQVYQLYTHNSNYQYNAQAGRL-----DLRQKAYLKLGN	290	
DB	111	NKEENKEAIGTLKTKLTDVTKDKIAKEQSKETIGTLAKTIDETVKDKIAKEQSKETI	170	
QY	291	DWABRLQELNIDVATRSQD-----TLEKV-----OERADK	322	
DB	171	GTLLKTLDETVKDKIAKEQSKETIGTLKTKLTDVTKDKIAKEQSKQDQKQALQKELAKK	230	

QY	308	SQTDLTLEK-----VQBRADKFEIENNTLKLN-----SDLSFNKALKOINDE	350
Db	567	KENELLOKKTILWTKTCSKEIALEQSELENERKRLKTTLSFKNLTQTLESLEKENSQ	626
QY	351	LTELSNAKEKRLKNDKSL-----SEKASKIQ-----ELEARK	383
QY	627	LDPE-----NLELRNRVSLKCSMKMAQLQLENKEJESK	662

3 RESULT

```

RESUL1      PCT-US01-04098A-1182
; Sequence 1182, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/593,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/554,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1182
; LENGTH: 990
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1182

```

Query Match 10.6%; Score 207.5; DB 1;
Best Local Similarity 23.0%; Pred. No. 4.7e-06;
Matches 106; Conservative 86; Mismatches 151;
Indels 117

Qy	7	RSQDTTLEKQVERADKFEIENNTLKLKNSDLSFNN---KALKDHNDLFTIELSNAKEUR	63
Db	296	RDELDALREKAVRVDKUESEVSRYSRKELHDIIEYKARVEELKEDNQVLLETMTMLEDD	355
Qy	64	GSAVTGRTIPDPQR-EALDKYELEHNDLKTNNEGLKTENEGLKTENEGSL--	120
Db	356	GT-----RARSOKLHELEKENQLKAKLHDMEEMDRDKKIEBEMEENTLEM	404
Qy	121	--KTEVDVRPRG-----TVENPOKA-----REL----LNKYDVENSMLQANN	158
Db	405	AQOKSMDESILHGWELEQISRTSELSEAPOSLSGHEFVELTSRLLKLEMNOSLTKTVE	464
Qy	159	KLPWRVRYTRHTPEDKCLKII-----DDIDAKEHELQQOONEKLSLONGDGHPREV	208
Db	465	EL-----RTTVDSVEGNASKILMKENQRLSKKVEILENEIVQ---EKQSJQNQNLSKOL	518
Qy	209	IEDLAANPAIGNIR-----LRHENK-----DLKARLENAMEVAGROFKRAGTL	252
Db	519	MKEAQLEKTIETLRENSERQIKILEQEENEHLNQTVSLSRQSORISAEARVKIEKENKI	578
Qy	253	LQOVQLTYKHNSNYQOYNQAAGRLL-----LRQAEYKLGKINDWABRLQELINIVATR	307
Db	579	L-----HESKTESYTSKUSKLTIEFKERQIKLEHYHKEGERAEELENELH----	625

[illegible]

2. RESULT

```

PCUS01-04098A-3150
; Sequence 3150, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 3150
; LENGTH: 931
; TYPE: PRT
; ORGANISM: Homo sapiens
PCUS01-04098A-3150

```

Query Match	10.6%;	Score	207.5;	DB 1;	Length	931;
Best Local Similarity	23.0%;	Pred. NO.	4.3e-06;			
Matches 106;	Conservative	86;	Mismatches	151;	Indels	117

[illegible]

```

QY 308 SOTDILEK-----VOERADKFEIENNTLKN-----SDLSFNKALXDHND 350
Db 626 KENELQKKITNLTKTEKTEAEQENSELERENKRLKTLOSKNLTFOLESLEKNSQ 685
QY 351 LTEELSNAKEKLKNDKSL---SEKASKIQ---ELEARK 383
Db 686 LDEE-----NLELRNVESLCAKMKMAQOLENKKELESEK 721

RESULT 4
PCT-US00-35017A-1265
; Sequence 1265, Application PC/TUS0035017A
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: PCT/US00/35017A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1265
; LENGTH: 1879
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1879)
; OTHER INFORMATION: Xaa = any amino acid or other as shown in Table 3
PCT-US00-35017A-1265

```

```

Query Match          9.4%; Score 184; DB 1; Length 1879;
Best Local Similarity 20.4%; Pred. No. 0.00028;
Matches 106; Conservative 83; Mismatches 171; Indels 160; Gaps 20;

QY      6 TRSDTGLE-KVQERADKFEIENNTLTKLKNSDLSFN-----NKALKDHNDELTEELS 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      948 SROELEKLRKLEGGDSDFHEQIADLQAOIAELAEKWLAKKEELQAAALARLDDEIAQK-N 1006
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      57 NAEKURGSVAVTGTINDPQR-----AKBALDQIELEHNDLTKTNEGKLYENE-----GLKT 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1007 NALKKIR---ELEIGHTSDLOEDLSERAARKAEKQKARDLGELEALKTELEDTLDSTAT 1063
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      109 ENGLKTENEGKLTVDVFRFGTVENPDNDKARELLNKY-----D 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1064 QQE-LRAKREQEVTVLKALNEETRSHEAQVQEMROKHAQAVQSLTEQLEQXKRAKANLD 1122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      148 VENSMLQANDKLPRVVR---YTRHTPEDKLKLT-----IDDLDAKEH 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1123 KKNQTTLEKENTDLAGELRVLQGAQOEVEHEMKKLQAVQELQSKSDGERARAELNDRVH 1182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      188 ELQOQNEKLS--LQNGDGNPREVTDLAANNPAIQNI-----RLRH 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1183 KLQNEVESVTGMLNBAEGKAIKLAKQVASSLOQDTQELLQEESROKLNIVSTSLRQLEE 1242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      227 ENKDLKARLENAMAYAG-----RDFKAGTLLDQVTVQLTYYTKNSN 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1243 ERNSLQDQDDEEMAKQLNHERHISTLNIQLSDSKKLQDFASTVBALEKKRFQKEIEN 1302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      267 Y-QQYNAQAGRLDRQKAEYKLGINDWAERLLQELN---IDVATRSQ-TDTLEKRVQBRAD 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1303 LTQQTTEKAAAYDKLEKT-----NRLQOELDLVDLDNQRLVSNLEKKQKQKFD 1353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      322 KFEIENNTLKLKNSD-----LSFNKALKDHP-- 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1354 QLLAAEKNISSKYADRDVREAEAREKETKALSARALEEALEAKEELERTNKMKLAEWG 1413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      348 -----NDELTEELSNAKELKLR--NDKSLSEKASKIOEL 379
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      1414  RFGSASKDVQGLSHDLEKSKRALGDPRLKMTQLEEL  1453

RESULT      5
PCT-US01-04098A-1169
; Sequence 1169, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1169
; LENGTH: 1047
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1169

```

	Query Match	9.1%	Score 177.5;	DB 1;	Length 1047;
	Best Local Similarity	22.3%	Pred: No. 0.00034;		
	Matches 94;	Conservative 62;	Mismatches 149;	Indels 117;	Gaps 15;
QY	13	LEKVOERADRFETENTLKLKNSDLSPFNKAL-----KDND-----ELTEE-LSNAK	59		
Db	497	LTTVAEKSRVLQLEEE-LTLRRGEITEELQOCLLHSGPPDPDPAAEILRLRRLLSABK	555		
QY	60	EKLRSASVTRGTINDPQRAKEALDYELENDHDKYNEGKLTNEGKLTNEGKLTNEGKLTNEG	119		
Db	556	EHQRESVLR-----DKYEKALKAYAEVDKMLRAANEKAYAEVAGLKDQVQOATSENNG	609		
QY	120	LKTEVDRVPRGTVPENPOKARELLNKYDVNSMLQANDKLPRVRYTRHTPEDKLKII	179		
Db	610	LM-----	618		
QY	180	DDLDAKEHELQOONKEKLSLQNGDGNPREVIDLAANNPAIQNLRHE-----NKDLKARL	235		
Db	619	DSL-ASDHQKSLIEDLKATINSGPAQKEIGELKA-----VMGIEKMEHQLCIGLQAKHDL	674		
QY	236	ENAMEYAGRDFKR-----AGTLDDQVTL---YTKHNSNTQOYNQAQGRDLDR--	280		
Db	675	ETAMHVKKEALREKLQAEALQGRHWRALQVQASHLELQEAQDQRRDAELRVH	734		
QY	281	-----KQAEYLKGLNDWAERLLQELINDVATRSOTDTLEKVOERADRFETEN	327		
Db	735	ELEKLDVEYRGOQAIEFLKEQISLAEKKM--LDYERLQRAEQAQKEQVESIRREKLLVAE	792		
QY	328	NTLKLKNSDLSPFNKALKDHNDLTELTSNAKE-----KLKRNKDSLSEKASKIOELE	380		
Db	793	NRLQAVEALCSQSHTHMIESND-ISEETIRKETVEGLQDKLNRDKDKEVTALTQSTEMLR	851		
QY	381	AR 382			

Db 852 AQ 853

```

RESULT      6
PCT-US01-04098A-3137
; Sequence 3137, Application PC/TUS0104098A
; GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
;   FILE REFERENCE: 21372-029
;   CURRENT APPLICATION NUMBER: PCT/US01/04098A
;   CURRENT FILING DATE: 2001-02-05
;   PRIOR APPLICATION NUMBER: Not yet Assigned
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: 09/728,422
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: 09/693,325
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 09/663,561
;   PRIOR FILING DATE: 2000-09-15
;   PRIOR APPLICATION NUMBER: 09/654,936
;   PRIOR FILING DATE: 2000-09-01
;   PRIOR APPLICATION NUMBER: 09/620,325
;   PRIOR FILING DATE: 2000-07-19
;   PRIOR APPLICATION NUMBER: 09/598,075
;   PRIOR FILING DATE: 2000-06-20
;   PRIOR APPLICATION NUMBER: 09/560,875
;   PRIOR FILING DATE: 2000-04-27
;   PRIOR APPLICATION NUMBER: 09/496,914
;   PRIOR FILING DATE: 2000-02-03
;   NUMBER OF SEQ ID NOS: 3960
;   SOFTWARE: Custom
;   SEQ ID NO 3137
;   LENGTH: 1047
;   TYPE: PRT
;   ORGANISM: Homo sapiens
PCT-US01-04098A-3137

```

```

Query Match          9.18; Score 177.5; DB 1; Length 1047;
Best Local Similarity 22.38; Pred. No. 0.00034;
Matches 94; Conservative 62; Mismatches 149; Indels 117; Gaps 15;

QY 13 LEKVOERADKFEIENNTLKNLSLSPFNKAL-----KDND-----ELTEE-LSNAK 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 497 LTTVAEKSRVLOEEB-LTLRGEIEELQOCLLHSGPPDPDPAEILRLRRLLSASK 555
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 EKLRGSAVTRGTINDPORAKEALDYELENDHLTKNEGLKTENEGLKTEGKTKTENEG 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 556 EHORESGVLR-----DKYEKALKAYQAEVDKLRANAENKAYQAEVAGLKVQOATSNMG 609
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 LKTEVDVRPFRGTVENPDKARELLKNKYDVENSLOANNDKLPWRVRYTRTPDEKCLKXII 179
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 610 LM-----DNMKS 618
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 DDLDAKEHELOQNEKLSLQONGDGNPREVIEDLAANPAIONIRLHE-----NKDLKARL 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 619 DSL-ASDHQKSLDLKATLNSGPGAQKEIGELKA--VMIEGIMKEHQLEIGNLOAKHDL 674
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 ENAMEVAGRDEFR-----AGTLLDQVTQL--YTKHNSNYQQVNAQAGRLDR-- 280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 ETAMHVKEKALREKLEQAEELAGLQRIHWRAQLEVOQASQHRLELQEAQDQDRDAELRVH 734
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 281 -----QKAEYLGKLNDAERLLQELINDIVATRSQDTLLEKVOERADKFIEN 327
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 735 ELEKLDVEYRQQAIEBFLKEQISLAEKKM--LDYERLQRAEQAGQKEVESIRLEKLVAE 792
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 NTLKLNLSLSPFNKALKDHNDELTEELSNAKE-----KLKRNKDLSLSKASKIOELE 380
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 793 NRLQAVEALCSQTHMIESND-ISEETIRKTEVBEGLQDKLNKRDKEVTALQSOTEMLR 851
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 AR 382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

D'b 852 AQ 853

```

RESULT
PCT-US01-10484-145
; Sequence 145, Application PC/TUS0110484
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE OF INVENTION: 21272-048
; CURRENT APPLICATION NUMBER: PCT/US01/10484
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/668,680
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/695,618
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/729,711
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: NOT YET ASSIGNED
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: Custom
; SEQ ID NO 145
; LENGTH: 2816
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-10484-145

```

```

Query Match      9.0%; Score 176; DB 1; Length 2816;
Best Local Similarity 22.7%; Pred. No. 0.0015;
Matches 105; Conservative 65; Mismatches 121; Indels 172; Gaps 23;

QY 9 QDTLEKVOERADKFEIENNTLKKN--SDLSFN---NKALKDHNDLDT---EELSNAK 59
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2105 ETEVEK-----NHIREVKSLLSELSFQKELNVQISERKTQTLTIKQIEKEE 2154
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 60 EKLRSASVTRGTINDPORAEALDKYELNHDLDKTKNEGLKTENEGLKTE-----N 110
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2155 ENLQ--VVLROMSKHKHTELKNILDMQLQENHEV---OGLKQHQDQSVSELSEKTVAVLE 2208
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 111 EGLKTEN-----EGLKTEVDYRFFP-----RGTVENPDKARE--LLNK 145
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2209 EKLEENLQISOQKGEIEWQKOLLERDKREIERTWAESRALQSCVECLSEKEDLQEK 2268
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 146 YDV-----ENS--MLQANDKDLPWRYRTHPTEDKLKTIIDDLDAKEHE 188
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2269 CDIWEKLAQTKRVLAASSENSKMEQSNLEKLEINVR-----KLQOELDQLN----- 2315
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 189 LQONEKLSLQONGDGNPREVIEDL-----AANPAIONIRLRHENKDLKARLENAM 239
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2316 ----RDKLSLNDHISAMQOQLQDVDTGLALTYSKDGCFSSPHLFLPKDLL----- 2364
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 240 EVAGDFKRACTLDQVTLQYTKHNSNYQYNAOAGRLDLRQAERYLKGUNDNAERLLQE 299
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2365 ----HTTKHDVLLSEQTRLQ-----KDISEWANR--FED 2393
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 300 LNIADVATRSQDITLEKVOERADKFEIENNTLKKNSDLSE-----NNKA 343
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2394 CQKEETKQQ--QLQVLQN-----EIEENKLUKVOEQEMFORLOKERESESKLETSKVT 2446
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 344 LKDHNDLSELSNAKEKLKNDKSLS-----EKASKIQELE 380
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

RESULT 8
PCT-US01-04098A-1516
: Sequence 1516, Application PC/TUS0104098A


```

; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1926
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1926

```

```

Query Match      8.9%; Score 173; DB 1; Length 1203;
Best Local Similarity 21.4%; Pred. No. 0.00077;
Matches 89; Conservative 61; Mismatches 142; Indels 124; Gaps 13;

```

```

QY 9 QDTTLEKVOE-----RADKFEI-----NNTLKNKNSDLSFNKALKDHNDELTE 53
DB 815 EQOTLNRALEEGKQREVLRGRGAELDEQKRLDRTVDLNLKELEKIGESKQALQLOA 874
QY 54 ELSNAKEKLRGSAV-----TRGTI-----NDPQRAKEALDKYELNHDLT 94
DB 875 QLEDYKEKARREVDARQAKDWASAEKTSGLRQDEIQRUQALQASQAEERTARL 934
QY 95 KNEGLKTENEGKLTENEGKLTENEGKLTENEGKLTENEGKLTENEGKLTENEGKLT 154
DB 935 DRELLAQRLQGLQEOAEAKKRSOD-----DRARLKG-----LE 968
QY 155 ANNDKLPWRVRYTRHTPE-----DKLKIITDDDAKEHELOQONEKLSLQNGDGNPREVIED 211
DB 969 EKYSRLETDELDEKNTVELLTDVRNRCQVDQLRTLMQE-----RSARQD 1015
QY 212 LAANNPAIONTELHNKDKLKALENAMVAGNDFRAGTLLDQVTLQYTKHNSYQOYN 271
DB 1016 LECD-----KISLRQNKDLTRLAS-----SEGQKPSASLSQLE 1051
QY 272 AQGRLDLRQAE-----YKGLNDWAERLLQELNIDVATRSQDTTLEKVOERADKFEI 326
DB 1052 SQQLQERLQAEEREKTVLQSTNRKLRKVKELSIQI-----EDRQHVNDQKQ----- 1102
QY 327 NNTLKNKNSDLSFNKALKDHNDELTELSNAKEKLRNKDKSLSEKASKTQLEAR 382
DB 1103 -----LSRLVKALKRQVDAAEEETERLDGLRKKRAQREVEEQHEVNEQLQAR 1148

```

```

RESULT 11
PCT-US01-04098A-1647
; Sequence 1647, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20

```

```

; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1647
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1647

```

```

Query Match      8.7%; Score 170.5; DB 1; Length 484;
Best Local Similarity 21.6%; Pred. No. 0.00035;
Matches 91; Conservative 73; Mismatches 134; Indels 123; Gaps 16;

```

```

QY 12 TLEKVOERADKFEIENNTLKNKNSDLSFNKALKDHNDELTELSNAKEKLRGSAVTRGT 71
DB 92 TIERLKEQQRDEREK-----QEEIDNVKKDLKLEKVV-----SLQGD 131
QY 72 INDPQRAKEALDKYELNHDLTKNNEGLKTENEGKLT-----ENEGKLTENEGKLT 123
DB 132 LSE-----KEA-SLLDLKEHASSLASSGLKKDSR-LKTLTALEQKKEECLKMESQLKAH 185
QY 124 VDRVFPRTGVNPKAREL---LNKYDVNSMLQANDKLPWRVRYTRHTPTPEKLLKIID 180
DB 186 EAALERARASPMDSRIQHLEREITRYKDESSAKAQAEVDRLLEILKEVENKDKDKKIAE 245
QY 181 DLDAAKEHELOQONEKLSLQNGDGNPREVIEDLAANNPAIONIRLHNKDKLKALENAM 240
DB 246 LESLTSROVKDKNKVA-----NLKHKEQVEKKKSQAMLE 280
QY 241 VAGRDFRAGTLLDQVTLQYTKHNSYQOYNQAQGRLDLRQAEYKGLND-----W 292
DB 281 EARR---REDNLNDSQLOD-----SLRKKDDRIEELREALRESVQIT 321
QY 293 AER---LLOELNIDVATRSQDTTLEKVOERADKFEIENNTLKNKNSDLSFNKAL----- 344
DB 322 AERWVLAQESARTNAEKQVSELLMAEKV-KQELSNKAKLSSTOOSLAKEKETHLTWL 380
QY 345 -----KDHNDLDELTE-----ELSNNAKEKLRNKDKSLS-EKASKIQEL 379
DB 381 RAERKHLVEVLEMKQBPALLAATSEKDNIALLELSSSKKTKTQEEVAALKREKDRLVQOL 440
QY 380 E 380
DB 441 K 441

```

```

RESULT 12
PCT-US01-04098A-3615
; Sequence 3615, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20

```


QY	64	-----GSVTRGTINDPQRAKALDKYELNHDLTKNNEGLKTENEGUKTEN	117
Db	396	YKVGAGSSLYGGTTINARKFEMNAELKENELA---QNRLELEKLRQDFEVTTON	452
QY	118	EGLTEVDVPRPTGVNP-----DKARELL-----NKYDV	148
Db	453	EKLVELRSAVEQVVKETPEYRCMQSFSVLYNESLQKAHLDEARTLLGTGTHQHV	512
QY	149	ENSMLOANNDKLPWRVYRTHPTEDKIKIIDDLDAKEHELOQ---ONEKL	196
Db	513	E---LIERDEVSLHKKLRETEVQLQEDTQAQYRKEVEMLRIEFQOTLAANEQAQPINREMRH	570
QY	197	---SLQNGD-----GNPREVIEDL-----AANPAQNIQLRH	226
Db	571	LISSLOHNHQLGVEVLYRKKLRQAQSDLNKTRLSGSSALLOSQSTEDPKDEPAELKP	630
QY	227	ENKDL-----KARLENAMEV---AGRDFKRAGTLLDQVTOVLYTKHNSYQ---QYNAQ	273
Db	631	DSEDLSOSSASKASQEDANEIKKAQESQKEMKLLLDWYRSAPKEQDKVQLMAAEKKS	690
QY	274	AGRLDLQKAEYKGLNDWAEERLQELNIDVATRSQDITLTKVQERADKFEIENNTLKLK	333
Db	691	AELEDLQRQ---LKDLEKKEKKMK-----ADEALKKIRAVEEQIEYQLKLAMA	740
QY	334	N-----SDLSFNKALKDHNDELTEELSNAKEKLRKNDKSLSEKASKIOEL	379
Db	741	KOEERALLSEMDVTGQAFEDMQEQNIRLMQQLREKDDANFKLMSIKSNQIHLK	793

RESULT 15
PCT-US01-10484-76
; Sequence 76, Application PC/TUS0110484
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-048
; CURRENT APPLICATION NUMBER: PCT/US01/10484
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/668,680
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 09/695,618
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/728,711
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: NOT YET ASSIGNED
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: Custom
; SEQ ID NO 76
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-10484-76

Query Match		7.9%	Score 155;	DB 1;	Length 456;
Best Local Similarity		20.7%	Pred. No. 0.0029;		
Matches		97;	Conservative	81;	Mismatches 159; Indels 132; Gaps 20;
QY	5	ATRSQDITLTKVQERADKFEIENNTLKNLSDSLSEFNKALKDHNDELTEELS	-----	56	
Db	17	ATASETFV---LQORMRIVEEQTSURDOLIMLDQEK---RSHVSIKERTSSVGLPSVI	72		
QY	57	-----NAKEKLGRS-----AVTRGTIND	74		
Db	73	PNSTRRVSPAPNLPMSKTSQDIGDSRISLTKTLNNAIKTMEGRLEGKIEILASRPLINDES	132		
QY	75	PQRAKALDKYELNHDLTKNNEGLKTENEGUKTEVEVDVPRPTGTV	134		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:40:13 ; Search time 44.17 Seconds
(without alignments)
1394.777 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 383

Sequence: 1 ACWVATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEARK 383

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1009251 seqs, 160854530 residues

Word size: 10

Total number of hits satisfying chosen parameters: 24

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database: Pending Patents_AA_Main:*

- 1: /cgn2.6/ptodata/2/paa/pctus_comb.pep.*
- 2: /cgn2.6/ptodata/2/paa/us06_comb.pep.*
- 3: /cgn2.6/ptodata/2/paa/us07_comb.pep.*
- 4: /cgn2.6/ptodata/2/paa/us080_comb.pep.*
- 5: /cgn2.6/ptodata/2/paa/us081_comb.pep.*
- 6: /cgn2.6/ptodata/2/paa/us082_comb.pep.*
- 7: /cgn2.6/ptodata/2/paa/us083_comb.pep.*
- 8: /cgn2.6/ptodata/2/paa/us084_comb.pep.*
- 9: /cgn2.6/ptodata/2/paa/us085_comb.pep.*
- 10: /cgn2.6/ptodata/2/paa/us086_comb.pep.*
- 11: /cgn2.6/ptodata/2/paa/us087_comb.pep.*
- 12: /cgn2.6/ptodata/2/paa/us088_comb.pep.*
- 13: /cgn2.6/ptodata/2/paa/us089_comb.pep.*
- 14: /cgn2.6/ptodata/2/paa/us090_comb.pep.*
- 15: /cgn2.6/ptodata/2/paa/us091_comb.pep.*
- 16: /cgn2.6/ptodata/2/paa/us092_comb.pep.*
- 17: /cgn2.6/ptodata/2/paa/us093_comb.pep.*
- 18: /cgn2.6/ptodata/2/paa/us094_comb.pep.*
- 19: /cgn2.6/ptodata/2/paa/us095_comb.pep.*
- 20: /cgn2.6/ptodata/2/paa/us096_comb.pep.*
- 21: /cgn2.6/ptodata/2/paa/us097_comb.pep.*
- 22: /cgn2.6/ptodata/2/paa/us098_comb.pep.*
- 23: /cgn2.6/ptodata/2/paa/us060_comb.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	383	100.0	383	15	US-09-151-409-16
2	50	13.1	443	7	US-08-325-278-6
3	35	9.1	440	17	US-09-302-756-35
4	33	8.6	42	1	PCT-US94-12268-31
5	33	8.6	42	5	US-08-143-412A-31
6	33	8.6	42	8	US-08-479-690-31
7	27	7.0	28	5	US-08-187-166-20
8	16	4.2	254	13	US-08-914-479-4
9	15	3.9	15	3	US-07-945-860B-4
10	15	3.9	15	3	US-07-945-860B-5

11 15 3.9 15 3 US-07-945-860B-6 Sequence 6, Appli
12 15 3.9 15 3 US-07-945-860B-7 Sequence 7, Appli
13 15 3.9 15 3 US-07-945-860B-9 Sequence 9, Appli
14 15 3.9 15 13 US-08-914-479-7 Sequence 7, Appli
15 15 3.9 15 13 US-08-914-479-8 Sequence 8, Appli
16 15 3.9 15 13 US-08-914-479-9 Sequence 9, Appli
17 15 3.9 15 13 US-08-914-479-10 Sequence 10, Appli
18 15 3.9 15 13 US-08-914-479-12 Sequence 12, Appli
19 15 3.9 281 13 US-08-914-479-6 Sequence 6, Appli
20 14 3.7 14 5 US-08-187-166-15 Sequence 15, Appli
21 13 3.4 13 5 US-08-187-166-23 Sequence 23, Appli
22 12 3.1 138 13 US-08-914-479-2 Sequence 2, Appli
23 11 2.9 12 5 US-08-187-166-16 Sequence 16, Appli
24 10 2.6 10 5 US-08-187-166-24 Sequence 24, Appli

ALIGNMENTS

RESULT 1
US-09-151-409-16
; Sequence 16, Application US/09151409A
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: GROUP A STREPTOCOCCAL VACCINES
; FILE REFERENCE: 48112.410
; CURRENT APPLICATION NUMBER: US/09/151.409A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,635
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Hexavalent M
; OTHER INFORMATION: fusion gene sequence constructed from
; OTHER INFORMATION: streptococcal type 24, 5, 6, 19, 1 and 3 M protein
; OTHER INFORMATION: DNAs
US-09-151-409-16

Query Match 100.0%; Score 383; DB 15; Length 383;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACWVATRSQDTLEKVOERADKFEIENNTLKNSDLSPFNKALKDHNDLTELSSNAKE 60
Db 1 ACWVATRSQDTLEKVOERADKFEIENNTLKNSDLSPFNKALKDHNDLTELSSNAKE 60
Qy 61 KLRSVATRGTTINDPQRAKALDKYELENHDLTKNKGKLTENEGLKTENEGLKTENEGL 120
Db 61 KLRSVATRGTTINDPQRAKALDKYELENHDLTKNKGKLTENEGLKTENEGLKTENEGL 120
Qy 121 KTEVDVFPFGTGVENPDKARELLNKYDVNSMLQANNDKLPWVRVYTRHTPEDKLLKIID 180
Db 121 KTEVDVFPFGTGVENPDKARELLNKYDVNSMLQANNDKLPWVRVYTRHTPEDKLLKIID 180
Qy 181 DLDAKEHELOQOQNEKLSLQNGDGNPREVIEDLAANPAIGNIRLRHENKDLKARLENAM 240
Db 181 DLDAKEHELOQOQNEKLSLQNGDGNPREVIEDLAANPAIGNIRLRHENKDLKARLENAM 240
Qy 241 VAGDFKFRAGTLLDOVTQLYTKINSYQYNAQAGRLDLRQKAEYKLGNDWAERLLQEL 300
Db 241 VAGDFKFRAGTLLDOVTQLYTKINSYQYNAQAGRLDLRQKAEYKLGNDWAERLLQEL 300
Qy 301 NIDVATRSQDTLEKVOERADKFEIENNTLKNSDLSPFNKALKDHNDLTELSSNAKE 360
Db 301 NIDVATRSQDTLEKVOERADKFEIENNTLKNSDLSPFNKALKDHNDLTELSSNAKE 360
Qy 361 KLRNKDKSLSEKASKIQELEARK 383

NAME/KEY:	Modified-site
LOCATION:	23
OTHER INFORMATION:	/note= "L1;I1;V1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	24
OTHER INFORMATION:	/note= "E2;Q1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	26
OTHER INFORMATION:	/note= "H1;I1;S1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	27
OTHER INFORMATION:	/note= "D1;R1;M1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	29
OTHER INFORMATION:	/note= "K1;R1;Q1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	30
OTHER INFORMATION:	/note= "T1;H1;A1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	31
OTHER INFORMATION:	/note= "K1;E1;N1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	33
OTHER INFORMATION:	/note= "E1;K1;D1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	34
OTHER INFORMATION:	/note= "G1;D1;K1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	36
OTHER INFORMATION:	/note= "K2;T1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	37
OTHER INFORMATION:	/note= "T2;A1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	38
OTHER INFORMATION:	/note= "E2;R1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	39
OTHER INFORMATION:	/note= "N2;L1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	40
OTHER INFORMATION:	/note= "E2;N1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	41
OTHER INFORMATION:	/note= "G1;N1;L1"
FEATURE:	
NAME/KEY:	Modified-site
LOCATION:	42
OTHER INFORMATION:	/note= "L2;A1"
PCT-US94-12268-31	

```
0; Gaps 0;
```

```
RESULT 5
US-08-143-412A-31
; Sequence 31, Application US/08143412A
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy J.
; APPLICANT: Ye, John
; APPLICANT: Kaminsky, Stephen M.
; APPLICANT: Hosein, Barbara H.
; APPLICANT: Nixon, Douglas F.
; APPLICANT: Koif, Wayne C.
; APPLICANT: Kowalski, Jacek
; TITLE OF INVENTION: Structured Synthetic Antigen Libraries
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Lisa Wilson
; STREET: 25 Davids Drive
; CITY: Hauppauge
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 11788
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/143,412A
; FILING DATE: 25-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, M. Lisa
; REGISTRATION NUMBER: 34,045
; REFERENCE/DOCKET NUMBER: 2008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-273-2828 x2119
; TELEFAX: 516-273-1717
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 2
; OTHER INFORMATION: /note= "T1;V1;*"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "V1;F1;*"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /note= "T1;N1;P1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5
; OTHER INFORMATION: /note= "R2;G1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 6
; OTHER INFORMATION: /note= "G2;D1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 7
; OTHER INFORMATION: /note= "T2;G1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 8
; OTHER INFORMATION: /note= "D1;R1;M1"
;
;
; OTHER INFORMATION: /note= "I1;N1;V1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: /note= "S1;P1;E1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 10
; OTHER INFORMATION: /note= "D1;R1;N1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 11
; OTHER INFORMATION: /note= "P2;E1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 12
; OTHER INFORMATION: /note= "Q1;V1;D1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /note= "R1;I1;K1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /note= "A2;E1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /note= "K1;D1;R1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 16
; OTHER INFORMATION: /note= "E2;L1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 17
; OTHER INFORMATION: /note= "A2;L1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 18
; OTHER INFORMATION: /note= "L2;A1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 19
; OTHER INFORMATION: /note= "D1;N2"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 20
; OTHER INFORMATION: /note= "K2;N1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 21
; OTHER INFORMATION: /note= "Y2;P1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 22
; OTHER INFORMATION: /note= "E1;A1;D1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "L1;I1;V1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 24
; OTHER INFORMATION: /note= "E2;Q1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "H1;I1;S1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "D1;R1;M1"
```

FEATURE:
NAME/KEY: Modified-site
LOCATION: 29
OTHER INFORMATION: /note= "K1;R1;Q1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 30
OTHER INFORMATION: /note= "T1;H1;A1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 31
OTHER INFORMATION: /note= "K1;E1;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 33
OTHER INFORMATION: /note= "E1;K1;D1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 34
OTHER INFORMATION: /note= "G1;D1;K1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 36
OTHER INFORMATION: /note= "K2;T1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 37
OTHER INFORMATION: /note= "T2;A1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 38
OTHER INFORMATION: /note= "E2;R1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 39
OTHER INFORMATION: /note= "N2;L1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 40
OTHER INFORMATION: /note= "E2;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 41
OTHER INFORMATION: /note= "G1;N1;L1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 42
OTHER INFORMATION: /note= "L2;A1"
US-08-143-412A-31

Query Match 8.6%; Score 33; DB 5; Length 42;
Best Local Similarity 100.0%; Pred. No. 4e-25;
Matches 33; ~ Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 74 DPQRAKALDYELNHDLTKEGLKTENEGL 106
Db 10 DPQRAKALDYELNHDLTKEGLKTENEGL 42

RESULT 6
US-08-479-690-31
Sequence 31, Application US/08479690
GENERAL INFORMATION:
APPLICANT: United Biomedical, Inc.
TITLE OF INVENTION: Structured Synthetic Antigen Libraries
TITLE OF INVENTION:
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin, Esq.
STREET: 345 Park Avenue
CITY: New York
STATE: New York

COUNTRY: U.S.A.
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,690
FILING DATE: 7-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/143,412
FILING DATE: 26-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lin, Maria C.H.
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4120US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2
OTHER INFORMATION: /note= "T1;V1;*"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 3
OTHER INFORMATION: /note= "V1;F1;*"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "T1;N1;P1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 5
OTHER INFORMATION: /note= "R2;G1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /note= "G2;D1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 7
OTHER INFORMATION: /note= "T2;G1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 8
OTHER INFORMATION: /note= "T1;N1;V1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /note= "S1;P1;E1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 10
OTHER INFORMATION: /note= "D1;R1;N1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 11
OTHER INFORMATION: /note= "P2;E1"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 12
OTHER INFORMATION: /note= "Q1;V1;D1"
FEATURE:
NAME/KEY: Modified-site

us-09-151-409-16.oligo.rapm

Tue May 15 07:28:35 2001

```

;
; LOCATION: 13
; OTHER INFORMATION: /note= "R1;I1;K1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /note= "A2;E1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /note= "K1;D1;R1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 16
; OTHER INFORMATION: /note= "E2;L1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 17
; OTHER INFORMATION: /note= "A2;L1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 18
; OTHER INFORMATION: /note= "L2;A1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 19
; OTHER INFORMATION: /note= "D1;N2"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 20
; OTHER INFORMATION: /note= "K2;N1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 21
; OTHER INFORMATION: /note= "Y2;P1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 22
; OTHER INFORMATION: /note= "E1;A1;D1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "L1;I1;V1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 24
; OTHER INFORMATION: /note= "E2;Q1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "H1;I1;S1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "D1;R1;M1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 29
; OTHER INFORMATION: /note= "K1;R1;Q1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 30
; OTHER INFORMATION: /note= "T1;H1;A1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 31
; OTHER INFORMATION: /note= "K1;E1;N1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 33
; OTHER INFORMATION: /note= "E1;K1;D1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 34
;
;
; OTHER INFORMATION: /note= "G1;D1;K1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 36
; OTHER INFORMATION: /note= "K2;T1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 37
; OTHER INFORMATION: /note= "T2;A1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 38
; OTHER INFORMATION: /note= "E2;R1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 39
; OTHER INFORMATION: /note= "N2;L1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 40
; OTHER INFORMATION: /note= "E2;N1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 41
; OTHER INFORMATION: /note= "G1;N1;L1"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 42
; OTHER INFORMATION: /note= "L2;A1"
;
; US-08-479-690-31
;
; Query Match 8.6%; Score 33; DB 8; Length 42;
; Best Local Similarity 100.0%; Pred. No. 4e-25;
; Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 74 DPQRAKALDKYELENHDLTKNEGLKTENEGL 106
; | | | | | | | | | | | | | | | | | | | | | |
; Db 10 DPQRAKALDKYELENHDLTKNEGLKTENEGL 42
;
; RESULT 7
; US-08-187-166-20
; ; Sequence 20, Application US/08187166
; ; GENERAL INFORMATION:
; ; APPLICANT: Klemm, Per
; ; APPLICANT: Sokurenko, Evgeni Veniaminovic
; ; APPLICANT: Hasty, David Long
; ; APPLICANT: Pallesen, Lars
; ; APPLICANT: Molin, Soren
; ; TITLE OF INVENTION: RECEPTOR SPECIFIC BACTERIAL ADHESINS AND
; ; TITLE OF INVENTION: THEIR USE
; ; NUMBER OF SEQUENCES: 55
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Browdy and Neimark
; ; STREET: 419 Seventh Street, N.W.
; ; CITY: Washington
; ; STATE: D.C.
; ; COUNTRY: U.S.A.
; ; ZIP: 20004
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Floppy disk
; ; COMPUTER: IBM PC compatible
; ; OPERATING SYSTEM: PC-DOS/MS-DOS
; ; SOFTWARE: Patent Release #1.0, Version #1.25
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/08/187,166
; ; FILING DATE: 27-JAN-1994
; ; CLASSIFICATION: 435
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Cooper, Iver P.
; ; REGISTRATION NUMBER: 28,005
; ; REFERENCE/DOCKET NUMBER: KLEMM-1
;

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 28 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-187-166-20

Query Match 7.0%; Score 27; DB 5; Length 28;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 KTKNEGLKTENEGKLTENEGKLTENEG 119
|||||
Db 1 KTKNEGLKTENEGKLTENEGKLTENEG 27

RESULT 8
US-08-914-479-4
Sequence 4, Application US/08914479
GENERAL INFORMATION:
APPLICANT: Dale, James B.
TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
FILE REFERENCE: 481112.404C2
CURRENT APPLICATION NUMBER: US/08/914,479
PRIOR FILING DATE: 1997-08-19
PRIOR APPLICATION NUMBER: 08/409,270
PRIOR FILING DATE: 1995-03-23
PRIOR FILING DATE: 1992-09-16
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 254
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An antigen of M5 and a carrier of the
OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479-4

Query Match 4.2%; Score 16; DB 13; Length 254;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 MVATRSQTDLEKVOE 18
|||||
Db 1 MVATRSQTDLEKVOE 16

RESULT 9
US-07-945-860B-4
Sequence 4, Application US/07945860B
GENERAL INFORMATION:
APPLICANT: Dale, James B.
TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
FILE REFERENCE: 481112.404C2
CURRENT APPLICATION NUMBER: US/07/945,860
PRIOR FILING DATE: 1997-08-19
PRIOR APPLICATION NUMBER: 08/409,270
PRIOR FILING DATE: 1995-03-23
PRIOR FILING DATE: 1992-09-16
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 254
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An antigen of M5 and a carrier of the
OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479-4

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,860B
FILING DATE: 16-SEP-1992
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 372.5770P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8394
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-945-860B-4

Query Match 3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 VATRSQTDLEKVOE 18
|||||
Db 1 VATRSQTDLEKVOE 15

RESULT 10
US-07-945-860B-5
Sequence 5, Application US/07945860B
GENERAL INFORMATION:
APPLICANT: Dale, James B.
TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
FILE REFERENCE: 481112.404C2
CURRENT APPLICATION NUMBER: US/07/945,860
PRIOR FILING DATE: 1997-08-19
PRIOR APPLICATION NUMBER: 08/409,270
PRIOR FILING DATE: 1995-03-23
PRIOR FILING DATE: 1992-09-16
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 254
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An antigen of M5 and a carrier of the
OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479-4

Query Match 4.2%; Score 16; DB 13; Length 254;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 MVATRSQTDLEKVOE 18
|||||
Db 1 MVATRSQTDLEKVOE 16

RESULT 9
US-07-945-860B-4
Sequence 4, Application US/07945860B
GENERAL INFORMATION:
APPLICANT: Dale, James B.
TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
FILE REFERENCE: 481112.404C2
CURRENT APPLICATION NUMBER: US/07/945,860
PRIOR FILING DATE: 1997-08-19
PRIOR APPLICATION NUMBER: 08/409,270
PRIOR FILING DATE: 1995-03-23
PRIOR FILING DATE: 1992-09-16
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4
LENGTH: 254
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An antigen of M5 and a carrier of the
OTHER INFORMATION: COOH-terminal portion of M5
US-08-914-479-4

us-09-151-409-16.oligo.rapm

Tue May 15 07:28:35 2001

```

Query Match          3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 AVTRGTINDPQRAKE 80
      |||||
DB 1  AVTRGTINDPQRAKE 15

RESULT 11
US-07-945-860B-6
; Sequence 6, Application US/07945860B
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; TITLE OF INVENTION: FOR GROUP A STREPTOCOCCAL VACCINE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,860B
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5770P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-945-860B-6

Query Match          3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRGTVPENPKAR 140
      |||||
DB 1  RVFPRGTVPENPKAR 15

RESULT 12
US-07-945-860B-7
; Sequence 7, Application US/07945860B
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; TITLE OF INVENTION: FOR GROUP A STREPTOCOCCAL VACCINE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,860B
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5770P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-945-860B-7

```

```

; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,860B
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5770P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-945-860B-7

Query Match          3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RVRYTRHTPEDKLLK 177
      |||||
DB 1  RVRYTRHTPEDKLLK 15

RESULT 13
US-07-945-860B-9
; Sequence 9, Application US/07945860B
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; TITLE OF INVENTION: FOR GROUP A STREPTOCOCCAL VACCINE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,860B
; FILING DATE: 16-SEP-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.5770P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-07-945-860B-9

```

Query Match 3.9%; Score 15; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 200 NGDGNPREVIEDLAA 214
| | | | | | | | | | | | | | | | | |
Db 1 NGDGNPREVIEDLAA 15

RESULT 14
US-08-914-479-7
; Sequence 7, Application US/08914479
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 481112.404C2
; CURRENT APPLICATION NUMBER: US/08/914,479
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; PRIOR FILING DATE: 1992-09-16
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: NH2-terminal fragment of M protein for
; OTHER INFORMATION: constructing antigens, which elicit opsonic
; OTHER INFORMATION: antibodies in an immunized animal
US-08-914-479-7

Query Match 3.9%; Score 15; DB 13; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 VATRSQTDTEKVKQE 18
| | | | | | | | | | | | | | | | | |
Db 1 VATRSQTDTEKVKQE 15

RESULT 15
US-08-914-479-8
; Sequence 8, Application US/08914479
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: ANTIGEN OF HYBRID M PROTEIN AND CARRIER
; FILE REFERENCE: 481112.404C2
; CURRENT APPLICATION NUMBER: US/08/914,479
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 08/409,270
; PRIOR FILING DATE: 1995-03-23
; PRIOR APPLICATION NUMBER: 07/945,860
; PRIOR FILING DATE: 1992-09-16
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: NH2-terminal fragment of M protein for
; OTHER INFORMATION: constructing antigens, which elicit opsonic
; OTHER INFORMATION: antibodies in an immunized animal
US-08-914-479-8

Query Match 3.9%; Score 15; DB 13; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 66 AVTRGTINDPQRAKE 80
| | | | | | | | | | | | | | | | | |
Db 1 AVTRGTINDPQRAKE 15

Search completed: May 10, 2001, 08:42:14
Job time: 121 sec

us-09-151-409-16.oligo.rapm

Tue May 15 07:28:35 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:40:28 ; Search time 6.16 Seconds
(without alignments)
424.431 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 383
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEARK 383

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 39611 seqs, 6826361 residues

Word size : 10
Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Pending Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	ID Description
1	35	9.1	440 4 US-08-302-756C-35 Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-08-302-756C-35
; Sequence 35, Application US/08302756C
; GENERAL INFORMATION:
; APPLICANT: FISCHETTI, Vincent A.
; APPLICANT: POZZI, Gianni
; APPLICANT: SCHNEEWIND, Olaf
; TITLE OF INVENTION: DELIVERY AND EXPRESSION OF A HYBRID SURFACE PROTEIN ON
; FILE OF INVENTION: THE SURFACE OF GRAM POSITIVE BACTERIA
; FILE REFERENCE: 016921-076
; CURRENT APPLICATION NUMBER: US/08/302,756C
; CURRENT FILING DATE: 1995-03-07
; PRIOR APPLICATION NUMBER: US 07/522,440
; PRIOR FILING DATE: 1990-05-11
; PRIOR APPLICATION NUMBER: US 07/742,199
; PRIOR FILING DATE: 1991-08-05
; PRIOR APPLICATION NUMBER: US 07/814,823
; PRIOR FILING DATE: 1991-12-23
; PRIOR APPLICATION NUMBER: US 07/851,082
; PRIOR FILING DATE: 1992-03-13
; PRIOR APPLICATION NUMBER: PCT/US93/02355

; PRIOR FILING DATE: 1993-03-12
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 440
; TYPE: PRT
; ORGANISM: S. pyogenes
US-08-302-756C-35

Query Match 9.1%; Score 35; DB 4; Length 440;
Best Local Similarity 100.0%; Pred. No. 6.2e-28;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 126 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 160
|||||
Db 1 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 35

Search completed: May 10, 2001, 08:42:23
Job time: 115 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:38:18 ; Search time 21.4 Seconds
(without alignments)
1023.062 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 383

Sequence: 1 ACWATRSQDTLEKVVQRA.....KNDKSLSEKASKIQLEARK 383

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 390729 seqs, 57163235 residues

Word size : 10

Total number of hits satisfying chosen parameters: 40

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

A_Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	383	100.0	383	20	Y04368
2	134	35.0	247	21	B03114
3	95	24.8	247	15	R50993
4	80	20.9	287	15	R50992
5	80	20.9	287	21	B03113
6	72	18.8	173	15	R50998
7	72	18.8	173	21	B03120
8	72	18.8	247	15	R50997
9	72	18.8	247	21	B03119
10	72	18.8	274	15	R50994
11	72	18.8	274	21	B03115
					Streptococcal hexa
					S. pyogenes hybrid
					Recombinant M24-M5
					Recombinant M24-M5
					S. pyogenes hybrid
					Recombinant M24-M5
					S. pyogenes hybrid
					Recombinant M19-M6
					S. pyogenes hybrid
					Recombinant M24-M5
					S. pyogenes hybrid

12	72	18.8	343	15	R50999	Recombinant tetra
13	72	18.8	343	21	B03121	S. pyogenes hybrid
14	35	9.1	100	6	P50295	Sequence encoded b
15	35	9.1	441	10	P09055	M6 streptococcal p
16	35	9.1	441	14	R41780	Streptococcus pyog
17	35	9.1	483	18	W08927	Type-6 M-protein.
18	33	8.6	42	16	R74258	ssAL streptococcal
19	31	8.1	31	8	P70907	type 6 streptococc
20	27	7.0	28	16	R76756	E. coli FimH MFP c
21	27	7.0	28	16	R51000	M24-M5-M6-M19 mult
22	18	4.7	67	21	B03123	S. pyogenes hybrid
23	18	4.7	135	15	R51001	M24-M5-M6-M19-M3-M
24	18	4.7	135	21	B03124	S. pyogenes octava
25	18	4.7	187	15	R50995	Recombinant M24-M5
26	18	4.7	187	21	B03116	S. pyogenes hybrid
27	18	4.7	305	15	R50996	Recombinant M24-M5
28	18	4.7	305	21	B03117	S. pyogenes hybrid
29	16	4.2	234	15	R50228	Sequence of a frag
30	16	4.2	284	15	R50229	Sequence of fragme
31	15	3.9	15	15	R50233	Sequence of NH2 te
32	15	3.9	15	15	R50230	Sequence of NH2 te
33	13	3.4	13	16	R76751	E. coli FimH MFP c
34	13	3.4	13	16	R76759	E. coli FimH MFP c
35	12	3.1	15	15	R50231	Sequence of LT-B-M
36	12	3.1	138	15	R50227	Pep M5 peptide eli
37	11	2.9	12	11	R04956	E. coli FimH MFP c
38	11	2.9	12	16	R76752	Pep M5 peptide eli
39	11	2.9	13	11	R04952	E. coli FimH MFP c
40	10	2.6	10	16	R76760	E. coli FimH MFP c

ALIGNMENTS

RESULT 1

Y04368
ID Y04368 standard; Protein; 383 AA.

XX
AC Y04368;

XX
DT 23-JUN-1999 (first entry)

XX
DE Streptococcal hexavalent M protein vaccine.

XX
DE Streptococcal; hexavalent emm gene; multivalent; vaccine; Group A;
KW immunogenic; immune response; pathogen; fusion protein.

XX
OS Streptococcus sp.

XX
PN W09913084-A1.

XX
PD 18-MAR-1999.

XX
PF 14-SEP-1998; 98WO-US19100.

XX
PR 12-SEP-1997; 97US-0058635.

XX
PA (IDVA-) ID VACCINE.

XX
PI Dale JB;

XX
DR WPI; 1999-215066/18.

XX
PT N-PSDB; X33103.

XX
PS Immunogenic fusion protein derived from group A streptococci

XX
CC Example 1; Fig 7; 50pp; English.

CC
CC The present invention describes an immunogenic fusion polypeptide (I)
CC that stimulates an immune response against group A streptococci (GAS).
CC (I) comprises: (a) at least 2 peptides (Ia) from GAS, at least 10 amino
CC acids long and able to induce a specific immune response; and (b) a
CC peptide (Ib), C-terminal to (Ia) that protects the immunogenicity of the

CC component described in (a) but is not essential for stimulation of the
 CC immune response. Vaccines containing (I) are used to protect against GAS
 CC infections, specifically those caused by Streptococcus pyogenes, e.g.
 CC pharyngitis, pyoderma, toxic shock syndrome, deep tissue invasion, sepsis
 CC and acute rheumatic fever. (I) have improved immunogenicity and do not
 CC generate antibodies that are cross-reactive with human tissues. The
 CC present sequence represents a hexavalent M protein vaccine.
 XX
 SQ Sequence 383 AA;

Query Match 100.0%; Score 383; DB 20; Length 383;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACWATRSQDTLEKVOERADKFEIENNTLKNSDLSFNKALKDHNDLTTELSNAKE 60
 DB 1 acwatrsqdtlekvoeradkfeientlknslsfnnkalkdhndelteelsnake 60
 QY 61 KLRSVAVTRGTINDPQRAKALDKYELNHDLTKNNEGLKTENEGKLTENEG 120
 DB 61 klrsavtrgtindpqrakealdkyelenhdltkneglkteneglkteneg 120
 QY 121 KTEVDVFPRTGVENPKARELLNKYDVENSMLQANNDKLPWRVRYTRHTPEDK 180
 DB 121 ktevdvfpgrtvenpdkarellnkydvensmlqanndklpwrvrytrhtpedk 180
 QY 181 DLDAKEHELOQNEKLSIQNGDNPREVIDEALANPAIGNIRLHNENKDKARLEN 240
 DB 181 dldakeheiqgnekslsiqngdnprevideealannpaignirlnhenkdka 240
 QY 241 VAGRDFKRAAGTLDDQVTLQYTKHNSNQYQNAQAGRLDLRQAEYKLGNDWAERLLQEL 300
 DB 241 vagrdfkragtlldqvvtlqytkhnsnyqqynaqagrldlrqaeylkgln 300
 QY 301 NIDVATRSQDTLEKVOERADKFEIENNTLKNSDLSFNKALKDHNDLTTELSNAKE 360
 DB 301 nidvatsqdtlekvoeradkfeientlknslsfnnkalkdhndelteelsnake 360
 QY 361 KLRKNDKLSKASKIOLEARK 383
 DB 361 klrknkslskaskioleark 383

RESULT 2
 B03114
 ID B03114 standard; Protein; 247 AA.
 XX
 AC B03114;
 DT 10-OCT-2000 (first entry)
 XX
 DE S. pyogenes hybrid M protein (M24-M5-M6-M19), SEQ ID NO:4.
 XX
 KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia.
 XX
 OS Streptococcus pyogenes.
 OS Synthetic.
 XX
 PN US6063386-A.
 XX
 PD 16-MAY-2000.
 XX
 PF 15-SEP-1997; 97US-0937271.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX

PI Lederer JW, Dale JB;
 XX
 DR WPI: 2000-364475/31.
 DR N-PSDB; A57894.
 XX

New immunogenic recombinant hybrid M protein comprising amino-terminal
 peptide fragments of streptococcal M protein useful as vaccine against
 rheumatic fever and infections leading to rheumatic fever -

PS Disclosure: Fig 4A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
 CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
 CC Streptococcus pyogenes M proteins generated in the disclosure of the
 CC invention.

XX Sequence 247 AA;

Query Match 35.0%; Score 134; DB 21; Length 247;
 Best Local Similarity 100.0%; Pred. No. 3.7e-113;
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 GSAVTRGTINDPQRAKALDKYELNHDLTKNNEGLKTENEGKLTENEGKLTKE 123
 DB 114 gsavtrgtindpqrakealdkyelenhdltkneglkteneglkteneglkte 173
 QY 124 VDRVFPRTGVENPKARELLNKYDVENSMLQANNDKLPWRVRYTRHTPEDK 183
 DB 174 vdrvfpgrtvenpdkarellnkydvensmlqanndklpwrvrytrhtpedk 233
 QY 184 AKEHELOQNEKLS 197
 DB 234 akehelqgneks 247

RESULT 3

R50993
 ID R50993 standard; Protein; 247 AA.

XX R50993;

XX 02-NOV-1994 (first entry)

XX Recombinant M24-M5-M6-M19.

XX Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;

Db 2 vatrsqtdtlekvqradkfeienntklknsdlsfnkalkdhndelteelsnakeklr 61

QY 364 KNDKSLSEKASKIQELEAR 383
|||||

Db 62 kndkselsekaskiqelear 81

RESULT 5

B03113

ID B03113 standard; Protein; 287 AA.

XX

AC B03113;

XX

DT 10-OCT-2000 (first entry)

XX

DE S. pyogenes hybrid M protein (M24-M5-M6), SEQ ID NO:2.

XX

XX Multivalent hybrid M protein; group A streptococcus; serotype;

KW immunogenic; sero-specific antibody; streptococcal infection;

KW cross reactivity; vaccine; acute rheumatic fever; ARF;

KW rheumatic heart disease; streptococcal pharyngitis; strep throat;

KW pneumonia.

XX

OS Streptococcus pyogenes.

OS Synthetic.

XX

PN US6063386-A.

XX

PD 16-MAY-2000.

XX

PF 15-SEP-1997; 97US-0937271.

XX

PR 16-SEP-1992; 92US-0945954.

XX

XX (UYTE-) UNIV TENNESSEE RES CORP.

PA Lederer JW, Dale JB;

XX

PI WPI; 2000-364475/31.

XX

DR N-PSDB; A57893.

XX

PT New immunogenic recombinant hybrid M protein comprising amino-terminal

PT peptide fragments of streptococcal M protein useful as vaccine against

PT rheumatic fever and infections leading to rheumatic fever -

XX

PS Disclosure; Fig 1A-B; 62pp; English.

XX

CC The invention relates to multivalent immunogenic hybrid group A

CC streptococcal M proteins comprising N-terminal peptide fragments of M

CC proteins that elicit opsonic antibodies against multiple serotypes

CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies

CC generated using the hybrid proteins are against one or more M protein

CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.

CC The invention also encompasses a recombinant DNA molecule comprising a

CC nucleotide sequence that encodes a multivalent hybrid M protein; and a

CC method for immunising a mammal against streptococcal infections,

CC comprising administering an immunogenic multivalent hybrid M protein to

CC the mammal. The multivalent hybrid M proteins are useful for eliciting

CC opsonic or protective antibodies to the M proteins of different serotypes

CC of group A streptococci, and may therefore be used as vaccines to protect

CC against and control infection by type A streptococci. Type A streptococci

CC are not only responsible for streptococcal pharyngitis (strep throat),

CC forms of pneumonia and a condition resembling toxic shock, but are also

CC involved in the development of acute rheumatic fever (ARF) and rheumatic

CC heart disease. In a patient with ARF, antibodies formed during a group A

CC streptococcal infection are also cross-reactive with heart tissue, which

CC indicates that the streptococci and host tissue contain similar antigenic

CC motifs. The new multivalent vaccines are capable of raising sero-specific

CC antibodies against various serotypes of group A streptococci which are

CC not cross-reactive with human heart tissue. Sequences B03113-B03117,

CC B03119-B03121 and B03123-B03124 represent multivalent hybrid

CC Streptococcus pyogenes M proteins generated in the disclosure of the

CC invention.

XX Sequence 287 AA;

SQ

Query Match 20.9%; Score 80; DB 21; Length 287;

Best Local Similarity 100.0%; Pred. No. 2.le-64;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQTDtlekvqradkfeienntklknsdlsfnkalkdhndelteelsnakeklr 363
|||||

Db 2 vatrsqtdtlekvqradkfeienntklknsdlsfnkalkdhndelteelsnakeklr 61

QY 364 KNDKSLSEKASKIQELEAR 383
|||||

Db 62 kndkselsekaskiqelear 81

RESULT 6

R50998

ID R50998 standard; Protein; 173 AA.

XX

AC R50998;

XX

DT 02-NOV-1994 (first entry)

XX

DE Recombinant M24-M5 divalent hybrid.

XX

XX Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;

KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;

KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;

KW restriction enzyme site; multivalent M protein; immunisation; group A;

KW streptococci; rheumatic fever; rheumatic heart disease; humoral;

KW antibody; heart tissue; antigen; serotype; mucosal.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Peptide 2..115

FT /label= M24

FT Peptide 116..173

FT /label= M5

XX

PN W09406421-A.

XX

XX 31-MAR-1994.

XX

PD 15-SEP-1993; 93WO-US08703.

XX

PF 16-SEP-1992; 92US-0945954.

XX

XX (UYTE-) UNIV TENNESSEE RES CORP.

PA Dale JB, Lederer JW;

PI WPI; 1994-118122/14.

XX

DR N-PSDB; Q45220.

XX

PT New immunogenic hybrid proteins derives from streptococcal M proteins

PT - induces opsonic antibodies, for protective immunisation against

PT against multiple group A streptococci serotypes

XX

PS Disclosure; Fig 10; 67pp; English.

XX

CC The sequences given in R50992-1001 represent hybrid M proteins which

CC contain the M24-M5-M6 and/or M19 subunits. These multivalent

CC proteins were constructed using fragments of the 5' regions of emm

CC genes that were amplified by PCR, ligated in tandem and expressed in

CC PKK223.3. The amplified regions pref. encode protective and not

CC tissue-cross-reactive epitopes, which can then be linked into one

CC protein molecule. The recombinant hybrid protein may contain 113

CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and

CC 35 from M19. Each section is linked by 2 amino acids specified by

CC the respective restriction enzyme sites that were synthesised into

CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.

XX
 SQ Sequence 173 AA;

Query Match 18.8%; Score 72; DB 15; Length 173;
 Best Local Similarity 100.0%; Pred. No. 2.4e-57;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 VATRSQDTLEKVRADKFEIENNTLKLKNSDLSEFNKALKDHNDTELSNAKEKLR 363
 Db 2 vatrsqdtlekvradkfeienntklknsdlsfnkalkdhndteelsnakeklr 61
 Qy 364 KNDKSISEKASK 375
 Db 62 kndksisekask 73

RESULT 7
 B03120
 ID B03120 standard; Protein; 173 AA.
 AC B03120;

DT 10-OCT-2000 (first entry)

DE S. pyogenes hybrid M protein (M24-M5), SEQ ID NO:15.

KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARP;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia.

XX Streptococcus pyogenes.

OS Synthetic.

XX US6063386-A.

PD 16-MAY-2000.

XX 15-SEP-1997; 97US-0937271.

XX 16-SEP-1992; 92US-0945954.

XX (UYTE-) UNIV TENNESSEE RES CORP.

PI Lederer JW, Dale JB;

XX WPI; 2000-364475/31.

DR N-PSDB; A57899.

XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever

XX Disclosure; Fig 10A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a

CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
 CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
 CC Streptococcus pyogenes M proteins generated in the disclosure of the
 CC invention.

XX Sequence 173 AA;

Query Match 18.8%; Score 72; DB 21; Length 173;
 Best Local Similarity 100.0%; Pred. No. 2.4e-57;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 VATRSQDTLEKVRADKFEIENNTLKLKNSDLSEFNKALKDHNDTELSNAKEKLR 363
 Db 2 vatrsqdtlekvradkfeienntklknsdlsfnkalkdhndteelsnakeklr 61
 Qy 364 KNDKSISEKASK 375
 Db 62 kndksisekask 73

RESULT 8

R50997

ID R50997 standard; Protein; 247 AA.

XX R50997;

XX 02-NOV-1994 (first entry)

DE Recombinant M19-M6-M4-M24 tetraivalent hybrid.

XX Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal.

XX Synthetic.

XX Key Location/Qualifiers

FT Peptide 2..38

FT /label= M19

FT Peptide 39..75

FT /label= M6

FT Peptide 76..135

FT /label= M5

FT Peptide 136..247

FT /label= M24

XX WO9406421-A.

XX 31-MAR-1994.

XX 15-SEP-1993; 93WO-US08703.

XX 16-SEP-1992; 92US-0945954.

XX (UYTE-) UNIV TENNESSEE RES CORP.

Tue May 15 07:28:29 2001

us-09-151-409-16.oligo.rag

```

XX PI Dale JB, Lederer JW;
XX XX WPI; 1994-110122/14.
DR DR N-PSDB; Q45219.
XX XX
XX XX New immunogenic hybrid proteins derives from streptococcal M proteins
PT PT - induces opsonic antibodies, for protective immunisation against
PT PT against multiple group A streptococci serotypes
XX XX
XX PS Disclosure; Fig 9; 67pp; English.
XX XX
XX XX The sequences given in R50992-1001 represent hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC PKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
CC to an immunogenic carrier.
XX XX
XX SQ Sequence. 247 AA;

Query Match 18.8%; Score 72; DB 15; Length 247;
Best Local Similarity 100.0%; Pred. No. 3.1e-57;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSOTDLEKVRADKFEIENNTLKLKNSDLFSFNKALKDHNDLTELSELSNAKEKLR 363
DB 136 vatrsqtdtlekvrqadkfeienttlklknsdlfsfnkalkdhndelteelsnakeklr 195
|||||
QY 364 KNDKSLSEKASK 375
DB 196 kndkslsekask 207
|||||

RESULT 9
ID B03119 standard; Protein; 247 AA.
XX AC B03119;
XX DT 10-OCT-2000 (first entry)
XX DE S. pyogenes hybrid M protein (M19-M6-M5-M24), SEQ ID NO:13.
XX XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia.
XX XX Streptococcus pyogenes.
OS Synthetic.
XX XX US6063386-A.
XX XX 16-MAY-2000.
XX XX 15-SEP-1997; 97US-09337271.
XX XX 16-SEP-1992; 92US-0945954.
PR

```

```

XX PA (UYTE-) UNIV TENNESSEE RES CORP.
XX PI Lederer JW, Dale JB;
XX XX WPI; 2000-364475/31.
XX XX
XX XX New immunogenic recombinant hybrid M protein comprising amino-terminal
PT PT peptide fragments of streptococcal M protein useful as vaccine against
PT PT rheumatic fever and infections leading to rheumatic fever -
XX XX
XX PS Disclosure; Fig 9A-B; 62pp; English.
XX XX
XX XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
CC Streptococcus pyogenes M proteins generated in the disclosure of the
CC invention.
XX XX
XX SQ Sequence 247 AA;

Query Match 18.8%; Score 72; DB 21; Length 247;
Best Local Similarity 100.0%; Pred. No. 3.1e-57;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSOTDLEKVRADKFEIENNTLKLKNSDLFSFNKALKDHNDLTELSELSNAKEKLR 363
DB 136 vatrsqtdtlekvrqadkfeienttlklknsdlfsfnkalkdhndelteelsnakeklr 195
|||||
QY 364 KNDKSLSEKASK 375
DB 196 kndkslsekask 207
|||||

RESULT 10
ID R50994
ID R50994 standard; Protein; 274 AA.
XX AC R50994;
XX XX 02-NOV-1994 (first entry)
XX DE Recombinant M24-M5-M6-M19 (linker variant).
XX XX
XX XX Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
KW restriction enzyme site; multivalent M protein; immunisation; group A;
KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
KW antibody; heart tissue; antigen; serotype; mucosal.
XX XX

```


OS Synthetic.
 XX Key Location/Qualifiers
 FH Peptide 1..115
 FT Peptide /label= M24
 FT Peptide 116..124
 FT Peptide /label= Linker
 FT Peptide 125..184
 FT Peptide /label= M5
 FT Peptide 185..193
 FT Peptide /label= Linker
 FT Peptide 194..230
 FT Peptide /label= M6
 FT Peptide 231..239
 FT Peptide /label= Linker
 FT Peptide 240..274
 FT Peptide /label= M5
 XX WO9406421-A.
 PN 31-MAR-1994.
 XX 15-SEP-1993; 93WO-US08703.
 XX 16-SEP-1992; 92US-0945954.
 XX (UYTE-) UNIV TENNESSEE RES CORP.
 PA Dale JB, Lederer JW;
 PI WPI; 1994-118122/14.
 DR N-PSDB; 045216.

XX New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX Disclosure; Fig 6; 67pp; English.

XX The sequences given in R50992-1001 represent hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC PKK223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
 CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.
 XX Sequence 274 AA;

Query Match 18.8%; Score 72; DB 15; Length 274;
 Best Local Similarity 100.0%; Pred. No. 3.4e-57;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 304 VATRSQDTLEKVOERADFEIENNTLKNKSDLSFNKALKDHNDELTEELSNKAKLR 363
 DB 2 vatrsqdtlekvgeradfeienntlkksdlsfnkalkdndelteesnakeklr 61
 QY 364 KNDKSLSEKASK 375
 DB 62 knkdsalsekask 73

RESULT 11
 B03115
 ID B03115 standard; Protein; 274 AA.
 XX AC B03115;
 XX DT 10-OCT-2000 (first entry)
 XX DE S. pyogenes hybrid M protein (M24-M5-M6-M19), SEQ ID NO:6.
 XX KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 XX pneumonia.
 XX OS Streptococcus pyogenes.
 XX OS Synthetic.
 XX PN US6063386-A.
 XX PD 16-MAY-2000.
 XX PF 15-SEP-1997; 97US-0937271.
 XX PR 16-SEP-1992; 92US-0945954.
 XX (UYTE-) UNIV TENNESSEE RES CORP.
 XX Lederer JW, Dale JB;
 XX WPI: 2000-364475/31.
 XX N-PSDB; A57895.
 XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever -
 XX Disclosure; Fig 6A-B; 62pp; English.
 XX The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
 CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
 CC Streptococcus pyogenes M proteins generated in the disclosure of the
 CC invention.
 XX Sequence 274 AA;

Query Match 18.8%; Score 72; DB 21; Length 274;
 Best Local Similarity 100.0%; Pred. No. 3.4e-57;

	Matches	72;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	304	VATRSQDTLEKVGQRADKFFIENNTIKLNSDLSFNNKALKDHNDLTELSSNAKEKLR	363							
DB	2	vatsrqdtlekvqradkfienntiklksdlsfnnkalkdhndelteelsnakeklr	61							
QY	364	KNDKSLSEKASK	375							
DB	62	kndkslseask	73							
RESULT	12									
R50999										
ID	R50999	standard; Protein; 343 AA.								
AC	R50999;									
XX										
DT	02-NOV-1994	(first entry)								
DE		Recombinant tetraivalent-C repeat.								
XX										
KW	Primer:	PCR: amplify; polymerase chain reaction; construct; hybrid;								
KW	M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;									
KW	tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;									
KW	restriction enzyme site; multivalent M protein; immunisation; group A;									
KW	streptococci; rheumatic fever; rheumatic heart disease; humoral;									
KW	antibody; heart tissue; antigen; serotype; mucosal.									
OS	Synthetic.									
XX										
Key		Location/Qualifiers								
FT	Peptide	1..38								
FT		/label= M19								
FT	Peptide	39..75								
FT		/label= M6								
FT	Peptide	76..135								
FT		/label= M5								
FT	Peptide	136..249								
FT		/label= M24								
FT	Peptide	250..343								
FT		/label= C-terminal_repeats								
XX										
PN	W09406421-A.									
XX										
PD	31-MAR-1994.									
XX										
PF	15-SEP-1993;	93WO-US08703.								
XX										
PR	16-SEP-1992;	92US-0945954.								
XX										
PA	(UVTE-) UNIV TENNESSEE RES CORP.									
XX										
PI	Dale JB, Lederer JW;									
XX										
DR	WPI: 1994-118122/14.									
DR	N-PSDB: Q45221.									
XX										
PT	New immunogenic hybrid proteins derives from streptococcal M proteins									
PT	- induces opsonic antibodies, for protective immunisation against									
PT	against multiple group A streptococci serotypes									
XX										
PS	Disclosure; Fig 11; 67pp; English.									
XX										
CC	The sequences given in R50992-1001 represent hybrid M proteins which									
CC	contain the M24-M5-M6 and/or M19 subunits. These multivalent									
CC	proteins were constructed using fragments of the 5' regions of emm									
CC	genes that were amplified by PCR, ligated in tandem and expressed in									
CC	PKK223.3. The amplified regions pref. encode protective and not									
CC	tissue-cross-reactive epitopes, which can then be linked into one									
CC	protein molecule. The recombinant hybrid protein may contain 113									
CC	N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6									

method for immunising a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic heart disease. In a patient with ARF, antibodies formed during a group A streptococcal infection are also cross-reactive with heart tissue, which indicates that the streptococci and host tissue contain similar antigenic motifs. The new multivalent vaccines are capable of raising sero-specific antibodies against various serotypes of group A streptococci which are not cross-reactive with human heart tissue. Sequences B03113-B03117, B03119-B03121 and B03123-B03124 represent multivalent hybrid Streptococcus pyogenes M proteins generated in the disclosure of the invention.

SQ Sequence 343 AA;

Query Match 18.8%; Score 72; DB 21; Length 343;

Best Local Similarity 100.0%; Pred. No. 4.1e-57; Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDELTEELSSNAKEKLR 363

Db 136 vatrsqdtlekvoeradkfeienntlklnsdlsfnmkalkdhndelteeisnakeklr 195

Qy 364 KNDKSLSEKASK 375

Db 196 kndkslsekask 207

RESULT 14

ID P50295 standard; Protein; 100 AA.

XX P50295;

XX 01-DEC-1991 (first entry)

XX Sequence encoded by a portion of the emm6 gene that encodes the amino terminal end of the M6 protein.

XX Fibrillar surface molecule; rheumatic fever; glomerulonephritis; throat swabs; diagnosis.

XX Streptococcus pyogenes (Group A streptococcus) strain D471.

XX Key Location/Qualifiers

FT Protein 43..100

FT /label= N-terminal of M6

XX W08500832-A.

XX 28-FEB-1985.

XX 09-AUG-1984; 84WO-4001261.

XX 18-JUN-1984; 84US-0521716.

PR 10-AUG-1983; 83US-0521962.

XX (UYRO-) ROCKEFELLER UNIV.

PA (UYEM-) EMORY UNIV.

PA (UYEM-) EMORY UNIV.

XX Scott JR, Fischetti VA;

XX WPI; 1985-062291/10.

DR N-PSDB; N50341.

XX

PT Immunogenic streptococcal polypeptide prodn. - by recombinant DNA methods, useful as vaccines and as diagnostic probes

XX Example: Fig 4; 46pp; English.

XX The inventors claim a polypeptide having an immunoreactive and antigenic determinant of a Streptococcus pyogenes M protein produced by a unicellular organism, and a DNA sequence encoding it. Also claimed are purified DNA probes able to bind to the S.pyogenes gene which codes for M protein. The polypeptides can be used in vaccines.

XX Sequence 100 AA;

Query Match 9.1%; Score 35; DB 6; Length 100;

Best Local Similarity 100.0%; Pred. No. 3.6e-24;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 RVFPRGTVENPKARELLNKYDVENSMLQANNDKL 160

Db 43 rvfprgtvenpkarelknkydvensmlqanndkl 77

RESULT 15

P90955

ID P90955 standard; protein; 441 AA.

XX AC P90955;

XX 23-FEB-1990 (first entry)

XX M6 streptococcal protein.

XX Immunoglobulin.

XX Streptococcus group A.

XX Key Location/Qualifiers

FT Region 216..235

FT Region 248..269

FT Region 275..284

XX W08909064-A.

XX 05-OCT-1989.

XX PF 13-MAR-1989; 89WO-US01026.

XX PR 25-MAR-1988; 88US-0173380.

XX (UYRO) ROCKEFELLER UNIV.

XX Fischetti V;

XX WPI; 1989-309382/42.

XX Polypeptide(s) from streptococcal M protein - used to prepare vaccines for providing protection against streptococcal infection

XX Disclosure; Fig 1; 22pp; English.

XX The regions (pref. conjugated to a natural carrier, eg cholera toxin) above can elicit an secretory Ig response in a mammal. They are used in vaccines against streptococcal infection, and give protection to different serotypes.

XX Sequence 441 AA;

Query Match 9.1%; Score 35; DB 10; Length 441;

Best Local Similarity 100.0%; Pred. No. 1.2e-23;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:38:53 ; Search time 13.07 Seconds
(without alignments)
562.950 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 383
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEAR 383

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 185757 seqs, 19210857 residues

Word size : 10
Total number of hits satisfying chosen parameters: 11

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	134	35.0	247	3	US-08-937-271-4	Sequence 4, Appli
2	80	20.9	287	3	US-08-937-271-2	Sequence 2, Appli
3	72	18.8	173	3	US-08-937-271-15	Sequence 15, Appl
4	72	18.8	247	3	US-08-937-271-13	Sequence 13, Appl
5	72	18.8	274	3	US-08-937-271-6	Sequence 6, Appli
6	72	18.8	343	3	US-08-937-271-17	Sequence 17, Appl
7	50	13.1	443	2	US-08-795-475-6	Sequence 6, Appli
8	18	4.7	67	3	US-08-937-271-20	Sequence 20, Appl
9	18	4.7	135	3	US-08-937-271-22	Sequence 22, Appl
10	18	4.7	187	3	US-08-937-271-8	Sequence 8, Appli
11	18	4.7	305	3	US-08-937-271-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-937-271-4
; Sequence 4, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/937,271
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-937-271-4
Query Match 35.0%; Score 134; DB 3; Length 247;
Best Local Similarity 100.0%; Pred. No. 8.6e-123;
Matches 134; Conservative 0; Mismatches 0; Gaps 0;
Indels 0;
QY 64 GSAVTRGTINDPQRAKALDKYLENHDLTKNEGLKTENEGKLTENEGKLTKE 123
|||||
Db 114 GSAVTRGTINDPQRAKALDKYLENHDLTKNEGLKTENEGKLTENEGKLTKE 173
QY 124 VDRVFRGTVENPDKARELKNKYDVNSMLQANNDKLPWRVRYTRHTPEDKLIIDDL 183
|||||
Db 174 VDRVFRGTVENPDKARELKNKYDVNSMLQANNDKLPWRVRYTRHTPEDKLIIDDL 233
QY 184 AKHEHQOQNEKLS 197
|||||
Db 234 AKHEHQOQNEKLS 247
RESULT 2
US-08-937-271-2
; Sequence 2, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/937,271
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997

```

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-2

Query Match 20.9%; Score 80; DB 3; Length 287;
Best Local Similarity 100.0%; Pred. No. 3.6e-70;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 2 VATRSQDTLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 61

QY 364 KNDKSLSEKASKIQELEAR 383
DB 62 KNDKSLSEKASKIQELEAR 81

RESULT 3
US-08-937-271-15
; Sequence 15, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-15

Query Match 18.8%; Score 72; DB 3; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.9e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 2 VATRSQDTLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 61

QY 364 KNDKSLSEKASK 375
DB 196 KNDKSLSEKASK 207

RESULT 5
US-08-937-271-6
; Sequence 6, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
```

```

Best Local Similarity 100.0%; Pred. No. 1.4e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 2 VATRSQDTLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 61

QY 364 KNDKSLSEKASK 375
DB 62 KNDKSLSEKASK 73

RESULT 4
US-08-937-271-13
; Sequence 13, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-13

Query Match 18.8%; Score 72; DB 3; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.9e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 363
DB 136 VATRSQDTLEKQVQRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKLR 195

QY 364 KNDKSLSEKASK 375
DB 196 KNDKSLSEKASK 207

RESULT 5
US-08-937-271-6
; Sequence 6, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
```

; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 274 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-6

Query Match 18.8%; Score 72; DB 3; Length 274;
Best Local Similarity 100.0%; Pred. No. 2.1e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQTDLEKVOERADKFEIENNTLKKNKSDLSFNKALKDHNDLTELSSNAKEKLR 363
|||||
Db 2 VATRSQTDLEKVOERADKFEIENNTLKKNKSDLSFNKALKDHNDLTELSSNAKEKLR 61
|||||
QY 364 KNDKSLSEKASK 375
|||||
Db 62 KNDKSLSEKASK 73
|||||

RESULT 6
US-08-937-271-17
; Sequence 17, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271

; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-17

Query Match 18.8%; Score 72; DB 3; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.6e-62;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQTDLEKVOERADKFEIENNTLKKNKSDLSFNKALKDHNDLTELSSNAKEKLR 363
|||||
Db 136 VATRSQTDLEKVOERADKFEIENNTLKKNKSDLSFNKALKDHNDLTELSSNAKEKLR 195
|||||
QY 364 KNDKSLSEKASK 375
|||||
Db 196 KNDKSLSEKASK 207
|||||

RESULT 7
US-08-795-475-6
; Sequence 6, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L-AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McWaters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-475-6

Query Match 13.1%; Score 50; DB 2; Length 443;

```
Best Local Similarity 100.0%; Pred. No. 8.4e-41; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 200 NGDGNPREVIEDLAANNPAIQNIRLHNKDLKARLENAMVAGRDFKRA 249
Db 1 NGDGNPREVIEDLAANNPAIQNIRLHNKDLKARLENAMVAGRDFKRA 50

RESULT 8
US-08-937-271-20
; Sequence 20, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 67 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-20

Query Match 4.7%; Score 18; DB 3; Length 67;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 EVDVFPFGTGVNPKAR 140
Db 33 EVDVFPFGTGVNPKAR 50

RESULT 9
US-08-937-271-22
; Sequence 22, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
```

```
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-937-271-22

Query Match 4.7%; Score 18; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 4e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 EVDVFPFGTGVNPKAR 140
Db 33 EVDVFPFGTGVNPKAR 50

RESULT 10
US-08-937-271-8
; Sequence 8, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 187 amino acids
```


Search completed: May 10, 2001, 08:41:04
Job time: 131 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: May 10, 2001, 08:39:43 ; Search time 18.37 Seconds
(without alignments)
1432.819 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 383
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIOLEARK 383

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 198801 seqs, 68722935 residues
Word size : 10

Total number of hits satisfying chosen parameters: 14

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

PIR_67:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	20.9	539	2	A28549 M24 protein precurs
2	58	15.1	492	1	A28616 M5 protein precurs
3	50	13.1	197	1	MMSOMP M5 protein - Strep
4	50	13.1	227	2	S01260 M1 protein precurs
5	50	13.1	484	2	S35401 M1 protein precurs
6	50	13.1	484	2	S46489 M protein precurs
7	36	9.4	83	2	S61086 M protein precurs
8	35	9.1	98	2	S61079 M protein precurs
9	35	9.1	150	2	S60838 M protein - Strep
10	35	9.1	483	2	A26297 M protein type 19
11	24	6.3	96	2	S71519 M protein precurs
12	18	4.7	87	2	S60786 M protein pepM19
13	16	4.2	26	2	PL0027 M protein precurs
14	14	3.7	87	2	S61075 M protein precurs

ALIGNMENTS

RESULT 1
A28549
M24 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M24
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 26-Aug-1999
C:Accession: A28549; S60802
R:Now, A.R.: Beachey, E.H.: Burdett, V.
J. Bacteriol. 170, 676-684, 1988
A:Title: Molecular evolution of streptococcal M protein: cloning and nucleotide sequence

A:Reference number: A28549; MUID:88115166
A:Accession: A28549
A:Molecule type: DNA
A:Residues: 1-539 <MOU>
R:Cross-references: GB:M19031; NID:g153616; PIDN:AAA26874.1; PID:g153617
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60802
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: DNA
A:Residues: 30-89 <WHA>
C:Superfamily: M5 protein
C:Keywords: coiled coil; transmembrane protein

Query Match 20.9%; Score 80; DB 2; Length 539;
Best Local Similarity 100.0%; Pred. No. 6.8e-72;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 304 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALKDHNDLFEELSNAKEKLR 363
|||||
Db 43 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALKDHNDLFEELSNAKEKLR 102
|||||

Qy 364 KNDKSLSEKASKIOLEARK 383
|||||
Db 103 KNDKSLSEKASKIOLEARK 122
|||||

RESULT 2

A28616
M5 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M5
C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 26-Aug-1999
C:Accession: A28616; S60787
R:Miller, L.; Gray, L.; Beachey, E.; Kehoe, M.
J. Biol. Chem. 263, 5668-5673, 1988
A:Title: Antigenic variation among group A streptococcal M proteins. Nucleotide sequence
A:Reference number: A28616; MUID:88186881
A:Accession: A28616
A:Molecule type: DNA
A:Residues: 1-492 <MIL>
A:Cross-references: GB:M20374; NID:g153812; PIDN:AAA26976.1; PID:g153813
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60787
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: DNA
A:Residues: 30-89 <WHA>
C:Genetics:
A:Gene: smp5
C:Superfamily: M5 protein
C:Keywords: coiled coil; transmembrane protein
F:1-42/Domain: signal sequence #status predicted <SIG>
F:43-492/Product: M5 protein #status predicted <MAT>

Query Match 15.1%; Score 58; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 7.1e-50;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 AVTRGINDPQAKALDKYLELNDLTKNEGLKTENEGLKTENEGLKTE 123
|||||
Db 43 AVTRGINDPQAKALDKYLELNDLTKNEGLKTENEGLKTENEGLKTE 100
|||||

RESULT 3

MMSOMP
M5 protein - Streptococcus pyogenes (fragment)

Tue May 15 07:28:40 2001

```
C:Species: Streptococcus pyogenes
C:Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 16-Feb-1997
C:Accession: A03501
R:Manjula, B.N.; Acharya, A.S.; Mische, S.M.; Fairwell, T.; Fischetti, V.A.
J. Biol. Chem. 259, 3686-3693, 1984
A:Title: The complete amino acid sequence of a biologically active 197-residue fragment
A:Reference number: A03501; MUID:84162039
A:Accession: A03501
A:Molecule type: protein
A:Comment: Residues 1-197 <HAA>
A:Residues: 1-197 <HAA>
C:Superfamily: M5 protein
C:Keywords: cell wall; transmembrane protein; virulence

Query Match      13.1%; Score 50; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 3.1e-42;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 DPORAKEALDKYLENHDKTKNEGLKTENEGKLTENEGKLTKE 123
      |||||
Db 9 DPORAKEALDKYLENHDKTKNEGLKTENEGKLTENEGKLTKE 58

RESULT 4
S01260
M protein precursor - Streptococcus pyogenes (serotype M1) (fragment)
C:Species: Streptococcus pyogenes
C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
C:Accession: S01260; S00767
R:Hannes-Fritz, E.
submitted to the EMBL Data Library, June 1988
A:Reference number: S01260
A:Accession: S01260
A:Molecule type: DNA
A:Residues: 1-227 <HAN>
A:Cross-references: EMBL:X07933; NID:g47358; PIDN:CAA30757.1; PID:g47359
A:Experimental source: strain CS130
R:Haanes-Fritz, E.; Kraus, W.; Burdett, V.; Dale, J.B.; Beachey, E.H.; Cleary, P.
Nucleic Acids Res. 16, 4667-4677, 1988
A:Title: Comparison of the leader sequences of four group A streptococcal M protein genes
A:Reference number: S00767; MUID:88247768
A:Accession: S00767
A:Molecule type: DNA
A:Residues: 1-199 <HAA>
A:Cross-references: EMBL:X07860
A:Experimental source: strain CS130
C:Genetics:
C:Gene: emm1
C:Superfamily: M5 protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-227/Product: M protein #status predicted <MAT>

Query Match      13.1%; Score 50; DB 2; Length 227;
Best Local Similarity 100.0%; Pred. No. 3.6e-42;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIQNIRLRHKNKDLKARLENAMVAGRDFKRA 249
      |||||
Db 42 NGDGNPREVIEDLAANNPAIQNIRLRHKNKDLKARLENAMVAGRDFKRA 91

RESULT 5
S35401
M1 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
A:Variety: serotype M1
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 26-Aug-1999
C:Accession: S35401; S61074; S60784
R:Pdbielecki, A.
submitted to the EMBL Data Library, September 1991

A:Reference number: S35401
A:Accession: S35401
A:Molecule type: DNA
A:Residues: 1-484 <PDB>
A:Cross-references: EMBL:XG2131; NID:g311757; PIDN:CAA4062.1; PID:g311758
R:Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.M.; Kehoe, M.
submitted to the EMBL Data Library, July 1994
A:Description: Noncongruent relationships between variation in emm1 gene sequences and
A:Reference number: S61072
A:Accession: S61072
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 16-94 <WHA>
A:Cross-references: EMBL:U11940; NID:g533557; PIDN:AAA99556.1; PID:g533558
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60784
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 29-89 <WH2>
A:Cross-references: EMBL:U11940
C:Genetics:
C:Gene: emm1
C:Superfamily: M5 protein

Query Match      13.1%; Score 50; DB 2; Length 484;
Best Local Similarity 100.0%; Pred. No. 7.3e-42;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIQNIRLRHKNKDLKARLENAMVAGRDFKRA 249
      |||||
Db 42 NGDGNPREVIEDLAANNPAIQNIRLRHKNKDLKARLENAMVAGRDFKRA 91

RESULT 6
S46489
M1 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-May-1999
C:Accession: S46489; S46490
R:Akesson, P.; Schmidt, K.H.; Cooney, J.; Bjoerck, L.
Biochem. J. 300, 877-886, 1994
A:Title: M1 protein and protein H: IgGFC- and albumin-binding streptococcal surface P
A:Reference number: S46489; MUID:94280417
A:Accession: S46489
A:Molecule type: DNA
A:Residues: 1-484 <AKE>
A:Experimental source: strain 40/58, serotype M1
A:Accession: S46490
A:Molecule type: protein
A:Residues: 42-51 <AKW>
A:Experimental source: strain 40/58, serotype M1
C:Genetics:
C:Gene: emm1
C:Superfamily: M5 protein
C:Keywords: transmembrane protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-484/Product: M1 protein #status experimental <MAT>
F:459-477/Domain: transmembrane #status predicted <TM>

Query Match      13.1%; Score 50; DB 2; Length 484;
Best Local Similarity 100.0%; Pred. No. 7.3e-42;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIQNIRLRHKNKDLKARLENAMVAGRDFKRA 249
      |||||
Db 42 NGDGNPREVIEDLAANNPAIQNIRLRHKNKDLKARLENAMVAGRDFKRA 91
```

```

RESULT 7
S61086
M protein precursor - Streptococcus pyogenes (serotype M68) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M68
C:Date: 15-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 16-Feb-1997
C:Accession: S61086; S60825
R:Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
A:Description: Noncongruent relationships between variation in emml gene sequences and the
A:Reference number: S61072
A:Accession: S61086
A:Molecule type: DNA
A:Residues: 1-83 <WHA>
A:Cross-references: EMBL:U11997
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60825
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 18-43, 'A', 45-78 <WHW>
A:Cross-references: EMBL:U11997
C:Genetics:
A:Gene: emm68
C:Superfamily: M5 protein
F:1-30/Domain: signal sequence (fragment) #status predicted <SIG>
F:31-83/Product: M protein (fragment) #status predicted <MAT>

Query Match          9.4%; Score 36; DB 2; Length 83;
Best Local Similarity 100.0%; Pred. No. 1.5e-28;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 ANPAQIQLRLHKNKDLKARLENAMVAGRDFKRA 249
      |||||
Db 45 ANPAQIQLRLHKNKDLKARLENAMVAGRDFKRA 80

RESULT 8
S61079
M protein precursor - Streptococcus pyogenes (serotype M19) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M19
C:Date: 15-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 16-Feb-1997
C:Accession: S61079; S60799
R:Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
A:Description: Noncongruent relationships between variation in emml gene sequences and the
A:Reference number: S61072
A:Accession: S61079
A:Molecule type: DNA
A:Residues: 1-98 <WHA>
A:Cross-references: EMBL:U11959
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60799
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 18-77 <WHW>
A:Cross-references: EMBL:U11959
C:Genetics:
A:Gene: emm19
C:Superfamily: M5 protein
F:1-30/Domain: signal sequence (fragment) #status predicted <SIG>
F:31-98/Product: M protein (fragment) #status predicted <MAT>

Query Match          9.1%; Score 35; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 1.8e-27;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RVRYTRHTPEDKLIKIIDDLDAKEHELOQQNEKLS 197
      |||||
Db 31 RVRYTRHTPEDKLIKIIDDLDAKEHELOQQNEKLS 65

RESULT 9
S60838
M protein precursor - Streptococcus pyogenes (serotype M6) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M6
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C:Accession: S60838
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60838
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-150 <WHA>
A:Cross-references: EMBL:U11986; NID:G533649; PIDN:AAA99602.1; PID:G1235835
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Superfamily: M5 protein

Query Match          9.1%; Score 35; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.6e-27;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 160
      |||||
Db 30 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 64

RESULT 10
A26297
M6 protein - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 26-Aug-1999
C:Accession: A26297
R:Hollingshead, S.K.; Fischetti, V.A.; Scott, J.R.
J. Biol. Chem. 261, 1677-1686, 1986
A:Title: Complete nucleotide sequence of type 6 M protein of the group A streptococcus
A:Reference number: A26297; MUID:86111835
A:Accession: A26297
A:Molecule type: DNA
A:Residues: 1-483 <HOL>
A:Cross-references: GB:M11338; GB:M11415; NID:G153699; PIDN:AAA26920.1; PID:G153700
C:Genetics:
A:Gene: emm6
C:Superfamily: M5 protein
C:Keywords: coiled coil; transmembrane protein

Query Match          9.1%; Score 35; DB 2; Length 483;
Best Local Similarity 100.0%; Pred. No. 7.9e-27;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 160
      |||||
Db 43 RVFPRGTVENPDKARELLNKYDVNSMLQANNDKL 77

RESULT 11
S71519
M protein type 19 - Streptococcus pyogenes (fragment)
C:Species: Streptococcus pyogenes
C:Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 07-May-1999
C:Accession: S71519; S71515
R:Podbielski, A.; Meizer, B.; Luttkien, R.
submitted to the EMBL Data Library, November 1990

```

A:Description: Application of the polymerase chain reaction to study the M protein(-like)
A:Reference number: S71518
A:Accession: S71519
A:Molecule type: DNA
A:Residues: 1-96 <POD>
A:Cross-references: EMBL:X56609
A:Experimental source: strain J17D/70
R:Podbielski, A.; Melzer, B.; Lueticken, R.
Med. Microbiol. Immunol. 180, 213-227, 1991
A:Title: Application of the polymerase chain reaction to study the M protein(-like) gene
A:Reference number: S71514; MUID:92149493
A:Accession: S71515
A:Molecule type: DNA
A:Residues: 1-42 <POW>
A:Cross-references: EMBL:X56609
A:Experimental source: J17D/70
C:Genetics:
A:Gene: emm19
C:Superfamily: M5 protein

Query Match 6.3%; Score 24; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.8e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RVRVTRHTPEDKLLKIIDDLDAKE 186
|||||
Db 43 RVRVTRHTPEDKLLKIIDDLDAKE 66
|||||

RESULT 12
S60786
M protein precursor - Streptococcus pyogenes (serotype M3) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M3
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: S60786
R:Whammore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the pop
A:Reference number: S60784; MUID:95198537
A:Accession: S60786
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-87 <WHA>
A:Cross-references: EMBL:U11945; NID:g533567; PIDN:AAA99561.1; PID:g533568
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994

Query Match 4.7%; Score 18; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 AQAGRLDLRQAEYLKGL 289
|||||
Db 70 AQAGRLDLRQAEYLKGL 87
|||||

RESULT 13
PL0027
M protein pepM19 - Streptococcus sp. (fragment)
C:Species: Streptococcus sp.
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Feb-1997
C:Accession: PL0027
R:Bronze, M.S.; Beachey, E.H.; Dale, J.B.
J. Exp. Med. 167, 1849-1859, 1988
A:Title: Protective and heart-crossreactive epitopes located within the NH2 terminus of
A:Reference number: PL0027; MUID:88258373
A:Accession: PL0027
A:Molecule type: protein
A:Residues: 1-26 <BRO>
C:Superfamily: M5 protein
C:Keywords: virulence factor

Query Match 4.2%; Score 16; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 5.6e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 PEDKLLKIIDDLDAKE 186
|||||
Db 8 PEDKLLKIIDDLDAKE 23
|||||

RESULT 14
S61075
M protein precursor - Streptococcus pyogenes (serotype M31) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M31
C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 15-Oct-1999
C:Accession: S61075; S60809
R:Whammore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.
submitted to the EMBL Data Library, July 1994
A:Description: Noncongruent relationships between variation in emm gene sequences an
A:Reference number: S61072
A:Accession: S61075
A:Molecule type: DNA
A:Residues: 1-87 <WHA>
A:Cross-references: EMBL:U11943; NID:g533563; PIDN:AAA99559.1; PID:g1235811
R:Whammore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60809
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 9-64, 'P' <WHW>
A:Cross-references: EMBL:U11943
C:Genetics:
A:Gene: emm31
F;1-16/Domain: signal sequence (fragment) #status predicted <SIG>
F;17-87/Product: M protein (fragment) #status predicted <MAT>

Query Match 3.7%; Score 14; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 QKAEYLKGLNDWAE 294
|||||
Db 69 QKAEYLKGLNDWAE 82
|||||

Search completed: May 10, 2001, 08:41:26
Job time: 103 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:41:08 ; Search time 11.87 Seconds
(without alignments)
1105.295 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 363
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEAR 363

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 93435 seqs, 34255486 residues

Word size : 10

Total number of hits satisfying chosen parameters: 3

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	80	20.9	539	1 M24_STRPY	P12379 streptococc
2	58	15.1	492	1 M5_STRPY	P02977 streptococc
3	35	9.1	483	1 M6_STRPY	P08089 streptococc

ALIGNMENTS

RESULT 1

M24_STRPY	1	
ID M24_STRPY	STANDARD;	PRT; 539 AA.
AC P12379;		
DT 01-OCT-1989 (Rel. 12, Created)		
DT 01-OCT-1989 (Rel. 12, Last sequence update)		
DT 01-OCT-1996 (Rel. 34, Last annotation update)		
DE M PROTEIN, SEROTYPE 24 PRECURSOR.		
GN EMM24.		
OS Streptococcus pyogenes.		
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC Streptococcus.		
OX NCBI_TaxID=1314;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=A24 VAUGHN;		
RX MEDLINE=88115166; PubMed=3276665;		
RA Mow A.R., Beachey E.H., Burdett V.;		
RT "Molecular evolution of streptococcal M protein: cloning and		
RT nucleotide sequence of the type 24 M protein gene and relation to		
RT other genes of Streptococcus pyogenes.";		
RL J. Bacteriol. 170:676-684(1988).		
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES		
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF		
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO		

CC PHAGOCYTOSIS.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.

CC -1- SIMILARITY: TO OTHER M PROTEINS.

CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS

CC IN THE REGION OF THE MEMBRANE ANCHOR.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M19031; AAA26874.1; -.

DR PIR; A28549; A28549.

DR InterPro; IPR001899; -.

DR Pfam; PF00746; Gram_pos_anchor; 1.

DR PRINTS; PR00015; GP0SANCHOR.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.

KW Signal; Virulence; Cell wall; Transmembrane; Phagocytosis;

KW Duplication; Repeat; Antigen; Coiled coil.

FT SIGNAL 1 42

FT CHAIN 43 539

FT DOMAIN 43 514

FT TRANSMEM 515 534

FT DOMAIN 535 539

FT DOMAIN 118 301

FT REPEAT 118 152

FT REPEAT 153 187

FT REPEAT 188 222

FT REPEAT 223 257

FT REPEAT 258 292

FT REPEAT 293 301

FT DOMAIN 311 405

FT REPEAT 311 355

FT REPEAT 356 380

FT REPEAT 381 405

FT DOMAIN 468 504

FT DOMAIN 505 510

FT SEQUENCE 539 AA; 58804 MW; B03EDF3AC1E6E9C7 CRC64;

Query Match 20.9%; Score 80; DB 1; Length 539;

Best Local Similarity 100.0%; Pred. No. 6.2e-70;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 304 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALDHNDLTELSENAKEKLR 363

DB 43 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALDHNDLTELSENAKEKLR 102

OY 364 KNDKSLSEKASKIQELEAR 383

DB 103 KNDKSLSEKASKIQELEAR 122

RESULT 2

M5_STRPY

ID M5_STRPY

AC P02977;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE M PROTEIN, SEROTYPE 5 PRECURSOR.

GN EMM5 OR SMP5.

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1314;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88186881; PubMed=3281944;

RA Miller L., Gray L., Beachey E., Kehoe M.;
 RT "Antigenic variation among group A streptococcal M proteins.
 RT Nucleotide sequence of the serotype 5 M protein gene and its
 RT relationship with genes encoding types 6 and 24 M proteins.";
 RL J. Biol. Chem. 263:5668-5673(1988).
 RN [2]
 RN SEQUENCE OF 43-212 AND 238-250.
 RN MEDLINE=84162039; PubMed=6368549;
 RX Manjula B.N., Acharya A.S., Mische S.M., Fairwell T., Fischetti V.A.;
 RA "The complete amino acid sequence of a biologically active
 RT 197-residue fragment of M protein isolated from type 5 group A
 streptococci.";
 RL J. Biol. Chem. 259:3686-3693(1984).
 RN [1]
 RN FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
 CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
 CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
 CC PHAGOCYTOSIS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 CC -1- SIMILARITY: TO OTHER M PROTEINS.
 CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 CC IN THE REGION OF THE MEMBRANE ANCHOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M20374; AAA26976.1; -
 DR PIR: A03501; MMSOMP.
 DR PIR: A28616; A28616.
 DR InterPro: IPR001899; -
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR PRINTS: PR00015; GPOSANCHOR.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
 DR VIRULENCE; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
 KW Transmembrane; Coiled coil; Signal.
 FT SIGNAL 1 42
 FT CHAIN 43 492 M PROTEIN, SEROTYPE 5.
 FT DOMAIN 43 466 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 467 486 MEMBRANE ANCHOR.
 FT DOMAIN 487 492 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 421 457 GLY/PRO-RICH (CELL WALL-SPANNING).
 FT DOMAIN 458 463 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 FT DOMAIN 69 103 5 X 7 AA TANDEM REPEATS OF L-K-T-K-N-E-G.
 FT REPEAT 69 75 1.
 FT REPEAT 76 82 2.
 FT REPEAT 83 89 3.
 FT REPEAT 90 96 4.
 FT REPEAT 97 103 5.
 FT CONFLICT 43 43 A -> T (IN REF. 2).
 FT CONFLICT 50 50 N -> S (IN REF. 2).
 FT CONFLICT 102 102 K -> SNLERKTAELTSEK (IN REF. 2).
 FT CONFLICT 208 208 I -> L (IN REF. 2).
 FT SEQUENCE 492 AA; 55085 MW; 873779B6CBD55E27 CRC64;
 SQ
 Query Match 15.1%; Score 58; DB 1; Length 492;
 Best Local Similarity 100.0%; Pred. No. 1.4e-48;
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 66 AVTRGTNDPQRAKEALDKYELNHDLTKNGLKTENEGLKTENEGLKTE 123
 DB 43 AVTRGTNDPQRAKEALDKYELNHDLTKNGLKTENEGLKTENEGLKTE 100
 RESULT 3
 M6_STRPY
 ID M6_STRPY PRT; 483 AA.
 AC P08089;

DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUL-1999 (Rel. 36, Last annotation update)
 DE M PROTEIN, SEROTYPE 6 PRECURSOR.
 GN EMM6.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus
 OX NCBI_TaxID=1314;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=86111835; PubMed=3511046;
 RA Hollingshead S.K., Fischetti V.F., Scott J.R.;
 RT "Complete nucleotide sequence of type 6 M protein of the group A
 streptococcus. Repetitive structure and membrane anchor.";
 RL J. Biol. Chem. 261:1677-1686(1986).
 RN [2]
 RN SEQUENCE OF 43-122 FROM N.A.
 RP MEDLINE=85166224; PubMed=3885219;
 RA Scott J.R., Pulliam W.M., Hollingshead S.K., Fischetti V.A.;
 RT "Relationship of M protein genes in group A streptococci.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:1822-1826(1985).
 RN [3]
 RN REVIEW.
 RP MEDLINE=91126460; PubMed=1846974;
 RX Fischetti V.A.;
 RA "Neural vector. Herpes may open the way to gene therapy in neurons.";
 RL Sci. Am. 264:32-39(1991).
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
 CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
 CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
 CC PHAGOCYTOSIS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 CC -1- SIMILARITY: TO OTHER M PROTEINS.
 CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 CC IN THE REGION OF THE MEMBRANE ANCHOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M11338; AAA26920.1; -
 DR PIR: A26297; A26297.
 DR InterPro: IPR001899; -
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR PRINTS: PR00015; GPOSANCHOR.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
 DR VIRULENCE; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
 KW Transmembrane; Coiled coil; Signal.
 FT SIGNAL 1 42 M PROTEIN, SEROTYPE 6.
 FT CHAIN 43 483 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 43 457 MEMBRANE ANCHOR.
 FT TRANSMEM 458 477 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 478 483 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 69 138 10 X 7 AA TANDEM REPEATS.
 FT DOMAIN 157 269 4.5 X 25 AA TANDEM REPEATS.
 FT DOMAIN 279 347 TWO DIRECTLY REPEATED 27 AMINO ACID
 FT BLOCKS SEPARATED BY 15 AMINO ACIDS.
 FT DOMAIN 348 411 HYDROPHILIC.
 FT DOMAIN 412 448 GLY/PRO-RICH (CELL WALL-SPANNING).
 FT DOMAIN 449 454 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 FT SEQUENCE 483 AA; 53472 MW; 68F87F28DB53A448 CRC64;
 SQ

Query Match 9.1%; Score 35; DB 1; Length 483;
 Best Local Similarity 100.0%; Pred. No. 3.2e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	126	RVEPRGTVPDKARELLNKYDVENSMQLQANDKL	160
Dd	43	RVEPRGTVPDKARELLNKYDVENSMQLQANDKL	77

Search completed: May 10, 2001, 08:43:09
Job time: 121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:40:53 ; Search time 27.21 Seconds
(without alignments)
1649.784 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 383
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEARK 383

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 374700 seqs, 117207915 residues

Word size : 10

Total number of hits satisfying chosen parameters: 24

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_15.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organelle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-unclassified.*
13: sp-vertebrate.*
14: sp-virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134	35.0	247	2 Q53344	Q53344 streptococc
2	52	13.6	457	2 Q54510	Q54510 streptococc
3	50	13.1	76	2 Q53536	Q53536 streptococc
4	50	13.1	79	2 Q54543	Q54543 streptococc
5	50	13.1	100	2 Q54632	Q54632 streptococc
6	50	13.1	272	2 Q9KGR9	Q9KGR9 streptococc
7	50	13.1	355	2 Q10371	Q10371 streptococc
8	50	13.1	484	2 Q10372	Q10372 streptococc
9	50	13.1	539	2 Q54719	Q54719 streptococc
10	45	11.7	100	2 Q54635	Q54635 streptococc
11	42	11.0	231	2 Q9KGR8	Q9KGR8 streptococc
12	38	9.9	100	2 Q54637	Q54637 streptococc
13	36	9.4	84	2 Q54598	Q54598 streptococc
14	35	9.1	99	2 Q54561	Q54561 streptococc
15	35	9.1	150	2 Q54587	Q54587 streptococc
16	30	7.8	488	2 Q54830	Q54830 streptococc
17	30	7.8	581	2 Q54835	Q54835 streptococc
18	26	6.8	100	2 Q54638	Q54638 streptococc
19	24	6.3	97	2 Q54845	Q54845 streptococc

20 23 6.0 83 2 086878 086878 streptococc
21 21 5.5 558 2 054718 054718 streptococc
22 19 5.0 83 2 Q92EH6 Q92EH6 streptococc
23 18 4.7 2 Q54548 Q54548 streptococc
24 14 3.7 87 2 Q54546 Q54546 streptococc

ALIGNMENTS

RESULT 1

Q53344
ID Q53344 PRELIMINARY; PRT; 247 AA.
AC Q53344;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE TETRAVALENT M PROTEIN (FRAGMENT).
GN EMM.
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX Streptococcus.
NX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93346765; PubMed=8345202;
RA Dale J.B., Chiang E.Y., Lederer J.W.;
RT "Recombinant tetraivalent group A streptococcal M protein vaccine.";
RL J. Immunol. 151:2188-2194(1993).
DR EMBL; S64396; AAB27745.1; -.
FT NON_TER 247
SQ SEQUENCE 247 AA; 28252 MW; 7AA5827DE945529B CRC64;

Query Match 35.0%; Score 134; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 7.2e-122;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 GSAVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTENEGLKTENEGLKTE 123
Db 114 GSAVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTENEGLKTENEGLKTE 173
QY 124 VDRVFPRTGVENPKARELLNKYDVENSMLQANDKLPWRVRYTRHTPEDKKKIIDDLD 183
Db 174 VDRVFPRTGVENPKARELLNKYDVENSMLQANDKLPWRVRYTRHTPEDKKKIIDDLD 233
QY 184 AKEHELOQOQNEKLS 197
Db 234 AKEHELOQOQNEKLS 247

RESULT 2

Q54510
ID Q54510 PRELIMINARY; PRT; 457 AA.
AC Q54510;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE M5.8193 PROTEIN (FRAGMENT).
GN EMM5.8193.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX Streptococcus.
NX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC8193;
RL MEDLINE=94224154; PubMed=8170398;
RA Whatmore A.M., Kehoe M.A.;
RT "Horizontal gene transfer in the evolution of group A streptococcal
emm-like genes: gene mosaics and variation in Vir regulons.";
RL Mol. Microbiol. 11:363-374(1994).
DR EMBL; U02480; AAA50854.1; -.

```
DR INTERPRO; IPR001899; -.
DR PFAM; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GP0SANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 457 AA; 51172 MW; DEBBAF550AEB6379 CRC64;

Query Match 13.6%; Score 52; DB 2; Length 457;
Best Local Similarity 100.0%; Pred. No. 4.8e-42; Indels 0; Gaps 0;
Matches 52; Conservative 0; Mismatches 0;

Qy 67 VTRGTINDPQRAEALDKYLENHDLTKKNEGLKTENEGKLTKTENEGKLTKTENE 118
Db 20 VTRGTINDPQRAEALDKYLENHDLTKKNEGLKTENEGKLTKTENEGKLTKTENE 71

RESULT 3
ID Q53536 PRELIMINARY; PRT; 76 AA.
AC Q53536;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE M PROTEIN (FRAGMENT).
GN EMM.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95377598; PubMed=7649435;
RA Penney T.J., Martin D.R., Williams L.C., de Malmanche S.A.,
RA Bergquist P.L.;
RT "A single emm gene-specific oligonucleotide probe does not recognise
RT all members of the Streptococcus pyogenes M type 1.";
RL FEMS Microbiol. Lett. 130:145-149(1995).
DR EMBL; S79254; AAB35162.1; -.
FT NON_TER 1
SQ SEQUENCE 76 AA; 8383 MW; EA46B64E5657A220 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 8.4e-41; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFFKRA 249
Db 22 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFFKRA 71

RESULT 4
ID Q54543 PRELIMINARY; PRT; 79 AA.
AC Q54543;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE M TYPE 1 (EMML) (FRAGMENT).
GN EMML.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 1.
RA Whatmore A.M.;
RL Thesis (1993), University of Newcastle Upon Tyne, UK.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 1.

RX MEDLINE=95198537; PubMed=7891551;
RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
RT "Non-congruent relationships between variation in emm gene sequences
RT and the population genetic structure of group A streptococci.";
RL Mol. Microbiol. 14:619-631(1994).
DR EMBL; U11940; AAA99556.1; -.
DR INTERPRO; IPR000886; -.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 79
SQ SEQUENCE 79 AA; 8499 MW; AE8A1FF8A4E1B711 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 8.7e-41; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFFKRA 249
Db 27 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFFKRA 76

RESULT 5
ID Q54632 PRELIMINARY; PRT; 100 AA.
AC Q54632;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)
DE M1.0 PROTEIN (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MGAS 2116;
RX MEDLINE=95172752; PubMed=7868273;
RA Musser J.M., Kapur V., Szezo J., Pan X., Swanson D.S., Martin D.R.;
RT "Genetic diversity and relationships among Streptococcus pyogenes
RT strains expressing serotype M1 protein: recent intercontinental spread
RT of a subclone causing episodes of invasive disease.";
RL Infect. Immun. 63:994-1003(1995).
DR EMBL; U20094; AAA85107.1; -.
FT NON_TER 1
FT NON_TER 100
SQ SEQUENCE 100 AA; 11275 MW; 63316019FC2B6CF2 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.1e-40; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

Qy 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFFKRA 249
Db 16 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFFKRA 65

RESULT 6
ID Q9KGR9 PRELIMINARY; PRT; 272 AA.
AC Q9KGR9;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE M PROTEIN (FRAGMENT).
GN EMM.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN=3148-00;
 RA Beall B.W.;
 RT #5' emm sequence emm3.2.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF288752; AAF91496.1;
 FT NON_TER 272 272
 FT NON_TER 272 272
 SQ SEQUENCE 272 AA; 30862 MW; 026FB243D345A35A CRC64;

Query Match 13.1%; Score 50; DB 2; Length 272;
 Best Local Similarity 100.0%; Pred. No. 2.6e-40;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 LLDQVTLTKHNSYQQYNAQAGRLDLKQAEYKGLNDWAERLQELN 301
 |||||
 Db 49 LLDQVTLTKHNSYQQYNAQAGRLDLKQAEYKGLNDWAERLQELN 98

RESULT 7
 Q10371 ID Q10371 PRELIMINARY; PRT; 355 AA.
 AC Q10371;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE M PROTEIN, SEROTYPE 1.0 PRECURSOR (FRAGMENTS).
 GN EMM1.0.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS130;
 RX MEDLINE=93360826; PubMed=8355619;
 RA Harbaugh M.P., Podbielski A., Huegl S., Cleary P.P.;
 RT "Nucleotide substitutions and small-scale insertion produce size and
 RT antigenic variation in group A streptococcal M1 protein.";
 RL Mol. Microbiol. 8:981-991(1993).
 RN [2]
 RP SEQUENCE OF 1-227 FROM N.A.
 RC STRAIN=CS130;
 RA Haanes-Fritz E., Kraus W., Burdett V., Dale J.B., Beachey E.H.,
 RA Cleary P.;
 RL Nucleic Acids Res. 16:4667-4676(1993).
 CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
 CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
 CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
 CC PHAGOCYTOSIS.
 CC -!- SURCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 DR EMBL: X07933; CAA30757.1;
 DR INTERPRO: IPR001899;
 DR PRINTS: PR00015; GPOSANCHOR.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
 KW Transmembrane; Coiled coil; Signal.
 FT SIGNAL 1 41
 FT CHAIN 42 355 M1.0 PROTEIN.
 FT NON_CONS 227 228
 FT DOMAIN 42 329 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 134 227 3 X APPROXIMATE TANDEM REPEATS, A-TYPE.
 FT DOMAIN 284 320 GLY/PRO-RICH.
 FT DOMAIN 321 325 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 FT TRANSMEM 330 349 MEMBRANE ANCHOR.
 FT DOMAIN 350 355 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 134 161 A-1.
 FT REPEAT 162 189 A-2.
 FT REPEAT 190 >227 A-3 (PARTIAL).
 SQ SEQUENCE 355 AA; 39595 MW; FF839CCAECDB01B9 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 3.3e-40;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFKRA 249
 |||||
 Db 42 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFKRA 91

RESULT 8
 Q10372 ID Q10372 PRELIMINARY; PRT; 484 AA.
 AC Q10372;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE M PROTEIN, SEROTYPE 1.2 PRECURSOR.
 GN EMM1.2.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ti/29/58;
 RX MEDLINE=93360826; PubMed=8355619;
 RA Harbaugh M.P., Podbielski A., Huegl S., Cleary P.P.;
 RT "Nucleotide substitutions and small-scale insertion produce size and
 RT antigenic variation in group A streptococcal M1 protein.";
 RL Mol. Microbiol. 8:981-991(1993).
 CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
 CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
 CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
 CC PHAGOCYTOSIS.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
 DR EMBL: X62131; CAA44062.1;
 DR INTERPRO: IPR001899;
 DR PRINTS: PF00746; Gram_pos_anchor; 1.
 DR PROSITE: PR00015; GPOSANCHOR.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 KW Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
 KW Transmembrane; Coiled coil; Signal.
 FT SIGNAL 1 41
 FT CHAIN 42 484 M1.2 PROTEIN.
 FT DOMAIN 42 458 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 134 244 3 X APPROXIMATE TANDEM REPEATS, A-TYPE.
 FT DOMAIN 245 379 3 X APPROXIMATE TANDEM REPEATS, C-TYPE.
 FT DOMAIN 413 449 GLY/PRO-RICH.
 FT DOMAIN 450 455 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 FT TRANSMEM 459 478 MEMBRANE ANCHOR.
 FT DOMAIN 479 484 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 134 161 A-1.
 FT REPEAT 162 189 A-2.
 FT REPEAT 190 244 A-3.
 FT REPEAT 245 286 C-1.
 FT REPEAT 287 327 C-2.
 FT REPEAT 328 379 C-3.
 SQ SEQUENCE 484 AA; 54234 MW; 4179A7BBAC5FAED9 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 484;
 Best Local Similarity 100.0%; Pred. No. 4.4e-40;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFKRA 249
 |||||
 Db 42 NGDGNPREVIEDLAANNPAIONIRLRHENKDLKARLENAMVAGRDFKRA 91

RESULT 9
 Q54719 ID Q54719 PRELIMINARY; PRT; 539 AA.

```
AC Q54719;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE M3 PROTEIN.
DE EMM3.1.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE 3;
RA Dale J.B., Simmons M., Chiang E., Chiang E.;
RL Vaccine 0:0-0(0).
DR EMBL; U40231; AAA96960.1; -.
DR HSSP; P03069; 1GCL.
DR INTERPRO; IPR001899; -.
DR PFAM; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GEOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 539 AA; 60163 MW; 93D20A4302C30F51 CRC64;

Query Match 13.1%; Score 50; DB 2; Length 539;
Best Local Similarity 100.0%; Pred. No. 4.8e-40; Indels 0; Gaps 0;
Matches 50; Conservative 0; Mismatches 0;

QY 252 LLDQVTLTKHNSYQYNAQAGRLDLRQKAEYKGLNDWAERLLQELN 301
Db 63 LLDQVTLTKHNSYQYNAQAGRLDLRQKAEYKGLNDWAERLLQELN 112

RESULT 10
Q54635 PRELIMINARY; PRT; 100 AA.
AC Q54635;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE M1.3 PROTEIN (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS 249;
RX MEDLINE=95172752; PubMed=7868273;
RA Musser J.M., Kapur V., Szeto J., Pan X., Swanson D.S., Martin D.R.;
RT "Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease.";
RL Infect. Immun. 63:994-1003(1995).
DR EMBL; U20097; AAA85110.1; -.
FT NON_TER 1
FT NON_TER 100
SQ SEQUENCE 100 AA; 11248 MW; 76FE8A3C33C15CF2 CRC64;

Query Match 11.7%; Score 45; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 7.7e-36; Indels 0; Gaps 0;
Matches 45; Conservative 0; Mismatches 0;

QY 205 PREVEDLAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRA 249
Db 21 PREVEDLAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRA 65

RESULT 11
Q9KGR8 PRELIMINARY; PRT; 231 AA.
ID Q9KGR8
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
AC Q9KGR8;
```

```
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE M PROTEIN (FRAGMENT).
DE GN EMM.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7203-99;
RA Beall B.W.;
RL "5' emm sequence emm3.3.";
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF288753; AAF91497.1; -.
FT NON_TER 1
FT NON_TER 231
SQ SEQUENCE 231 AA; 26305 MW; F5002430DD229351 CRC64;

Query Match 11.0%; Score 42; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.3e-32; Indels 0; Gaps 0;
Matches 42; Conservative 0; Mismatches 0;

QY 260 YTKHNSYQYNAQAGRLDLRQKAEYKGLNDWAERLLQELN 301
Db 58 YTKHNSYQYNAQAGRLDLRQKAEYKGLNDWAERLLQELN 99

RESULT 12
Q54637 PRELIMINARY; PRT; 100 AA.
ID Q54637;
AC Q54637;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE M1.5 PROTEIN (FRAGMENT).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS 1666;
RX MEDLINE=95172752; PubMed=7868273;
RA Musser J.M., Kapur V., Szeto J., Pan X., Swanson D.S., Martin D.R.;
RT "Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease.";
RL Infect. Immun. 63:994-1003(1995).
DR EMBL; U20099; AAA85112.1; -.
FT NON_TER 1
FT NON_TER 100
SQ SEQUENCE 100 AA; 11217 MW; 5332ACD6FFE7A3F2 CRC64;

Query Match 9.9%; Score 38; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 4.8e-29; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 0;

QY 212 LAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRA 249
Db 28 LAANNPAIQNIRLRHENKDLKARLENAMVAGRDFKRA 65

RESULT 13
Q54598 PRELIMINARY; PRT; 84 AA.
ID Q54598
AC Q54598;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
AC Q54598;
```



```
DE M TYPE 68 (EMML) (FRAGMENT).
GN EMML.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 68;
RA Whatmore A.M.;
RL Thesis (1993), University of Newcastle Upon Tyne, UK.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 68;
RX MEDLINE=95198537; PubMed=7891551;
RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
RT "Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci.";
RL Mol. Microbiol. 14:619-631(1994).
DR EMBL; U11997; AAA99613.1; -.
FT NON_TER 1
FT NON_TER 84
SQ SEQUENCE 84 AA; 9123 MW; BE68CCD5AF3C518F CRC64;

Query Match          9.4%; Score 36; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 3.6e-27;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 ANPAQNTLRHNRKDLKARLENAMVAGRDFKRA 249
Db 45 ANPAQNTLRHNRKDLKARLENAMVAGRDFKRA 80
|||||
|||||

RESULT 14
Q54561 PRELIMINARY; PRT; 99 AA.
AC Q54561;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE M TYPE 19 (EMML) (FRAGMENT).
GN EMML.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 19;
RA Whatmore A.M.;
RL Thesis (1993), University of Newcastle Upon Tyne, UK.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 19;
RX MEDLINE=95198537; PubMed=7891551;
RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
RT "Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci.";
RL Mol. Microbiol. 14:619-631(1994).
DR EMBL; U11959; AAA99575.1; -.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11034 MW; B04429D3DD2DF2E8 CRC64;

Query Match          9.1%; Score 35; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 3.8e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RVRYRHTPDKLLKIIDLDKAHEHQOQNEKLS 197
Db 31 RVRYRHTPDKLLKIIDLDKAHEHQOQNEKLS 65
|||||
|||||
```

```
RESULT 15
Q54587 PRELIMINARY; PRT; 150 AA.
AC Q54587;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE EMML PROTEIN (FRAGMENT).
GN EMML.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 6;
RA Whatmore A.M.;
RT "Sequence Analysis of the Emm-Like Gene Family of Streptococcus
Pyogenes.";
RL Thesis (1993), Microbiology, University of Newcastle-Upon-Tyne, UK.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-M TYPE 6;
RX MEDLINE=95198537; PubMed=7891551;
RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
RT "Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci.";
RL Mol. Microbiol. 14:619-631(1994).
DR EMBL; U11986; AAA99602.1; -.
FT NON_TER 1
FT NON_TER 150
SQ SEQUENCE 150 AA; 16572 MW; 8E00EAA28AC71157 CRC64;
```

Query Match 9.1%; Score 35; DB 2; Length 150;
Best Local Similarity 100.0%; Pred. No. 5.6e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RVFPRGTVPDKARELLNKYDVNSMLQANNDKL 160
Db 30 RVFPRGTVPDKARELLNKYDVNSMLQANNDKL 64
|||||

Search completed: May 10, 2001, 08:42:54
Job time: 121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:57:50 ; Search time 1527.63 Seconds
(without alignments)
11179.422 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcacgacgtgcgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1283235 seqs, 7373929652 residues

Word size : 10
Total number of hits satisfying chosen parameters: 988154

Minimum DB seq length: 7
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_in1.*
- 5: gb_in2.*
- 6: gb_in3.*
- 7: gb_om.*
- 8: gb_ov.*
- 9: gb_pat1.*
- 10: gb_pat2.*
- 11: gb_ph.*
- 12: gb_pl1.*
- 13: gb_pl2.*
- 14: gb_pl3.*
- 15: gb_pl4.*
- 16: em_bal.*
- 17: em_ba2.*
- 18: em_fun.*
- 19: em_htgo_hum.*
- 20: em_htgo_inv.*
- 21: em_htgo_rod.*
- 22: em_htg_hum1.*
- 23: em_htg_hum2.*
- 24: em_htg_hum3.*
- 25: em_htg_hum4.*
- 26: em_htg_hum5.*
- 27: em_htg_hum6.*
- 28: em_htg_hum7.*
- 29: em_htg_hum8.*
- 30: em_htg_inv1.*
- 31: em_htg_inv2.*
- 32: em_htg_other.*
- 33: em_htg_rod.*
- 34: em_hum1.*
- 35: em_hum2.*
- 36: em_hum3.*
- 37: em_hum4.*
- 38: em_hum5.*
- 39: em_hum6.*
- 40: em_hum7.*
- 41: em_in.*
- 42: em_om.*
- 43: em_or.*

- 44: em_ov.*
- 45: em_pat.*
- 46: em_ph.*
- 47: em_pl.*
- 48: em_ro.*
- 49: em_sts.*
- 50: em_sy.*
- 51: em_un.*
- 52: em_vi.*
- 53: gb_sts1.*
- 54: gb_sts2.*
- 55: gb_sts3.*
- 56: gb_sy.*
- 57: gb_un.*
- 58: gb_vil.*
- 59: gb_vl2.*
- 60: gb_vl1.*
- 61: gb_vl2.*
- 62: gb_vl3.*
- 63: gb_vl4.*
- 64: gb_vl5.*
- 65: gb_vl6.*
- 66: gb_vl7.*
- 67: gb_vl8.*
- 68: gb_vl9.*
- 69: gb_vl10.*
- 70: gb_vl11.*
- 71: gb_vl12.*
- 72: gb_vl13.*
- 73: gb_vl14.*
- 74: gb_vl15.*
- 75: gb_vl16.*
- 76: gb_vl17.*
- 77: gb_vl18.*
- 78: gb_vl19.*
- 79: gb_vl20.*
- 80: gb_vl21.*
- 81: gb_vl22.*
- 82: gb_vl23.*
- 83: gb_vl24.*
- 84: gb_vl25.*
- 85: gb_vl1.*
- 86: gb_vl2.*
- 87: gb_vl3.*
- 88: gb_vl4.*
- 89: gb_vl5.*
- 90: gb_vl6.*
- 91: gb_vl7.*
- 92: gb_vl8.*
- 93: gb_vl9.*
- 94: gb_vl10.*
- 95: gb_vl11.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	402	34.7	741	3	S64396 emm=tetrava
2	240	20.7	2131	3	M19031 S.pyogenes
3	174	15.0	1547	3	M20374 S.pyogenes
4	158	13.6	2633	3	U02480 Streptococc
5	151	13.0	229	3	S79254 emm=M prote
6	151	13.0	237	3	U11940 Streptococc
7	151	13.0	300	3	U20094 Streptococc
8	151	13.0	817	2	AF288752 Streptococ
9	151	13.0	1332	9	AR079002 Sequence
10	151	13.0	1494	3	X62131 S.pyogenes
11	151	13.0	1620	3	UA0231 Streptococc

596	17	1.5	88031	60	AC004378	AL132793	Human DNA
597	17	1.5	88528	85	AC003681	AL136987	Human DNA
598	17	1.5	88623	68	AC023565	AL512303	Homo sapi
599	17	1.5	88882	91	HS448E20	AC008364	Drosophila
600	17	1.5	89428	65	AC017242	AL513483	Homo sapi
601	17	1.5	90142	13	ATAC016827	AC008396	Homo sapi
602	17	1.5	93593	13	AF263283	AC008882	Homo sapi
603	17	1.5	93979	12	AC068564	AL136101	Human DNA
604	17	1.5	94818	12	AC006429	AC005558	Drosophila
605	17	1.5	95190	12	AC007203	AC015757	Homo sapi
606	17	1.5	95612	12	AC007661	AL020668	Human DNA
607	17	1.5	95893	61	AC009184	AL445228	Homo sapi
608	17	1.5	96059	13	AC006413	AL392174	Arabidopsis
609	17	1.5	96574	13	ATT18B16	AC008420	Homo sapi
610	17	1.5	97818	62	AF069716	AL049709	Human DNA
611	17	1.5	98123	12	AC027665	AL365442	Homo sapi
612	17	1.5	98231	12	AC027665	AC007726	Drosophila
613	17	1.5	98741	60	AC008578	AC008777	Homo sapi
614	17	1.5	99886	91	HS598F2	AL356782	Human DNA
615	17	1.5	100000	90	AP000037	AC078792	Homo sapi
616	17	1.5	100000	90	AP000105	AL136179	Human DNA
617	17	1.5	100000	90	AP000181	AC021265	Homo sapi
618	17	1.5	100000	90	AP000215	AC008835	Homo sapi
619	17	1.5	100389	3	SL100KBER	AC022103	Homo sapi
620	17	1.5	100469	13	ATF19H22	AC020761	Homo sapi
621	17	1.5	100749	66	AC020229	AC079830	Oryza sat
622	17	1.5	100953	87	AC009489	AL445204	Human DNA
623	17	1.5	101461	62	AC010658	AC006423	Arabidopsis
624	17	1.5	102512	62	AC016158	AL512975	Sulfolobu
625	17	1.5	102751	92	HS237H22	AL017755	Drosophila
626	17	1.5	103785	13	ATF1116	U75930	Orygia pseu
627	17	1.5	103796	61	AC010372	AL499610	Homo sapi
628	17	1.5	104278	61	AC009352	AP003023	Oryza sat
629	17	1.5	104436	91	HS611N7	AC004520	Homo sapi
630	17	1.5	104674	91	HS740A11	AC025716	Caenorhab
631	17	1.5	104817	91	HS681J21	AC003999	Human PAC
632	17	1.5	105308	13	AF273333	AC004600	Homo sapi
633	17	1.5	105413	61	AC009342	AC009892	Homo sapi
634	17	1.5	106704	60	AC008574	AP002861	Oryza sat
635	17	1.5	106748	92	HS832E24	AL022145	Homo sapi
636	17	1.5	106795	85	AC004864	AC069346	Homo sapi
637	17	1.5	106926	60	AC007691	AC010877	Homo sapi
638	17	1.5	108259	65	AC017510	AC022275	Homo sapi
639	17	1.5	108716	91	HS101AA	AC015605	Mus muscu
640	17	1.5	109343	86	AC007161	AC007712	Drosophila
641	17	1.5	109745	80	AC026447	AC026745	Homo sapi
642	17	1.5	109866	78	AF043945	AC002519	Human Chr
643	17	1.5	110000	75	AC073804_0	AC023266	Homo sapi
644	17	1.5	110000	83	CEY10666_1	AC084837	Homo sapi
645	17	1.5	110000	83	CEY57G11_1	AC012227	Homo sapi
646	17	1.5	110000	84	PFMAL13P2_0	AC008212	Drosophila
647	17	1.5	110000	83	CNS06C7K	AC002430	Human BAC
648	17	1.5	110754	14	IGS02P16	AC016448	Homo sapi
649	17	1.5	111117	80	AL358932	AC016448	Oryza sat
650	17	1.5	111118	92	HSAC000112	AC078893	Homo sapi
651	17	1.5	111126	88	AF321237	AC016837	Homo sapi
652	17	1.5	111135	89	AL135933	AC005177	Homo sapi
653	17	1.5	111464	14	F10K1	AC018778	Homo sapi
654	17	1.5	111945	13	ATF1C12	AC006404	Mus muscu
655	17	1.5	112418	60	AC008288	AC006404	Homo sapi
656	17	1.5	113215	82	AP001864	AC020629	Homo sapi
657	17	1.5	113252	85	AC004408	AP001550	Oryza sat
658	17	1.5	116878	89	AL135780	AC015996	Homo sapi
659	17	1.5	117296	12	AC008153	AC015996	Homo sapi
660	17	1.5	118603	60	AC008213	AP002871	Oryza sat
661	17	1.5	118684	85	AC004259	297192	Human DNA s
662	17	1.5	118995	85	AC005368	AP028899	Oryza sat
663	17	1.5	119111	13	ATF13C5	AC048359	Homo sapi
664	17	1.5	119407	13	ATFSK20	AC069251	Genomic s
665	17	1.5	119420	92	HS997K18	AC010347	Homo sapi
666	17	1.5	119927	85	AC005062	AC055823	Homo sapi
667	17	1.5	120171	60	AC007755	AC007258	Arabidopsis
668	17	1.5	120515	89	AL136230		


```
/partial
/feature="emmm"
/note="hybrid molecule containing amino-terminal subunits
of types 24, 5, 6, and 19 M proteins; This sequence comes
from Fig. 1"
/codon_start=1
/product="tetraivalent M protein"
/protein_id="AAB27745.1"
/db_xref="GI:408225"
/translation="MVATRSQDTLKVQRADKFEIENNTLKLKNSDLSPNNKALKD
HNDELTELSNAKEKLRLKNDKSLSEKASKNOLEAKADLEKALEGAMNFSTADSAKI
KTLAEKADLEKSAVTRGTTNDPORAKEALDKYLENHDLKTKNEGKLTENEGKLTEN
EGLTKENEGKLTVDYRVEPCTVENPKARELLNKYDVENSMLQANNKDLPLWRVRYTR
HTPDKLKKIIDLDAKEHELOQNEKLS"
BASE COUNT      311 a 106 c 164 g 160 t
ORIGIN
```

```
Query Match      34.7%; Score 402; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 3.6e-204;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 ggaatccgcgtgactaggggtacataaataagcccgcaagagcaaaagagctcttgac 249
    |||||||
Db 340 GCATCCGCGTCACTAGGGGTACAATAAATGACCCGCAAGAGCAAAAGAGCTCTTGAC 399

QY 250 aagatgagctagaaacacatgaactaaataaactaagaatgaagggttaaaactgagaat 309
    |||||||
Db 400 AAGTATGAGCTAGAAAACCATGACTTAAAACTAAGAAATGAAGGTTAAAACTGAGAT 459

QY 310 gaagggttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaactgag 369
    |||||||
Db 460 GAAGGTTAAAACTGAGAAATGAAGGTTAAAACTAAGAAATGAAGGTTAAAACTGAG 519

QY 370 gtcaacagaggttttcttaggggacggttagaaacccggacaaagcacgagaactctt 429
    |||||||
Db 520 GTCGACAGAGTGTTCCTATGGGGGACGGTAGAAACCCGGACAAAGCAGAGAACTTCTT 579

QY 430 aacaagtatgacgtagagaactctatgtacaagcttaataataacaaagttaccatggaga 489
    |||||||
Db 580 AACAAGTATGACGTAGAGAACTCTATGTTACAGCTAATAATGACAAAGTTACCATGGAGA 639

QY 490 gtgcgttactagctacgcagagaagataagctaaaaaattattgacgactcttgac 549
    |||||||
Db 640 GTGCGTTATATAGTCATAGCCAGAGAGATAAGCTAAAAAATTTATTGACGATCTTGAC 699

QY 550 gcaaaagaacatgaattacacacacagaaatgagaagtattct 591
    |||||||
Db 700 GCAAAAGAACATGAAATTACACACACAGAAATGAGAAGTTATCT 741
```

```
RESULT 2
STREMM24
LOCUS      STREMM24      2131 bp      DNA      BCT      26-APR-1993
DEFINITION S.pyogenes type 24 M-protein gene, complete cds.
ACCESSION  M19031
VERSION     M19031.1 GI:153616
KEYWORDS   type 24 M-protein.
SOURCE      S.pyogenes DNA, clone pVB41-L3.
ORGANISM   Streptococcus pyogenes
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE  1 (bases 1 to 2131)
AUTHORS   Mow,A.R., Beachey,E.H. and Burdett,V.
TITLE      Molecular evolution of Streptococcal M protein: Cloning and
            nucleotide sequence of the type 24 M protein gene and relation to
            other genes of Streptococcus pyogenes
JOURNAL    J. Bacteriol. 170, 676-684 (1988)
MEDLINE    88115166
COMMENT    Draft entry and computer-readable sequence for [1] kindly provided
            by V.Burdett, 21-APR-1988.
FEATURES   Location/Qualifiers
            source      1. .2131
```

```
/organism="Streptococcus pyogenes"
/db_xref="taxon:1314"
157. .1776
/note="type 24 M-protein precursor"
/codon_start=1
/transl_table=11
/protein_id="AAA26874.1"
/db_xref="GI:153617"
/translation="MTKNNTNRYSLRLKLTGTASVAVALTVLGAGLVVNTNEVSAVA
TSQDITLSEKQVQRADKFEIENNTLKLKNSDLSPNNKALKDHNDELTELSNAKEKLRL
KNDKSLSEKASKIOLEAKADLEKALEGAMNFSTADSAKI KTLAEKAAALAAKADL
EKALEGAMNFSTADSAKI KTLAEKAAALAAKADLEKALEGAMNFSTADSAKI KTLAE
EKAAALAAKADLEKALEGAMNFSTADSAKI KTLAEKAAALAAKADLEKALEGAMNF
TADSAKI KTLAEKAAALAAKADLEKALEGAMNFSTADSAKI KTLAEKAAALAAKAD
LEEQNKISEASRSQSLRRDLDAAREAKKOLEAHPQNKISEASRSQSLRRDLDAARE
QAEELAKLRAGKASDTPDAPGNKAVPGKGOAPQAGTKPNONKAPMKETKROLPS
GETANPFTAAALTVMATAGVAADVVRKEEN"
157. .282
sig_peptide
/note="pot. type 24 M-protein signal peptide; putative"
283. .1773
mat_peptide
/note="type 24 M-protein"
1802. .1838
misc_feature
/note="transcription termination signal; putative"
848 a 376 c 421 g 486 t
BASE COUNT
ORIGIN 100 bp upstream of HincII site.

Query Match      20.7%; Score 240; DB 3; Length 2131;
Best Local Similarity 100.0%; Pred. No. 2.4e-117;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaaaagatacaagaacgctgcacaaagttt 969
    |||||||
Db 283 GTCGCGACTAGTCTCAGACAGATATCTCGAAAAGTACAAAGACGCTGACAAAGTTT 342

QY 970 gagatagaaaaacaatcgttaaaacttaagaatagtgacttaagttaaataaagcg 1029
    |||||||
Db 343 GAGATAGAAAACAATACGTTAAACTTAAAGATAGTACTTAAGTTTAAATTAAGACG 402

QY 1030 ttaaaagatcataatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1089
    |||||||
Db 403 TTAAGAGATCAATATGATGAGTTAACTGAAGAGTTGAGTAATGCTTAAAGAGAACTAGT 462

QY 1090 aaaaatgataatcactatctgaaaaagctagtaaaattcaagaattagagcagcagtaag 1149
    |||||||
Db 463 AAAAATGATAATCACTATCTGAAAAGCTAGTAAATTCAGAAATTAGAGCACGTAAAG 522

RESULT 3
STREMP5A
LOCUS      STREMP5A      1547 bp      DNA      BCT      26-APR-1993
DEFINITION S.pyogenes smp5 gene encoding serotype 5 M protein, complete cds.
ACCESSION  M20374
VERSION     M20374.1 GI:153812
KEYWORDS   5 M protein; smp5 gene.
SOURCE      S.pyogenes (strain Manfredo) DNA, clone pMK207.
ORGANISM   Streptococcus pyogenes
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE  1 (bases 1 to 1547)
AUTHORS   Miller,L., Gray,L., Beachey,E.H. and Kehoe,M.
TITLE      Antigenic variation among group A streptococcal M proteins:
            Nucleotide sequence of the serotype 5 M protein gene and its
            relationship with genes encoding types 6 and 24 M proteins
JOURNAL    J. Biol. Chem. 263, 5668-5673 (1988)
MEDLINE    88186881
FEATURES   Location/Qualifiers
            source      1. .1547
            /organism="Streptococcus pyogenes"
            /db_xref="taxon:1314"
            43. .1521
            /note="5 M protein"
```



```

Qy 317 taaaactgagaatgaagggttaaaactgagaatgaa 354
|||||
Db 176 TAAAACTGAGATGAAGGTTAAAAAAGTGAATGAA 213

RESULT 5
S79254
LOCUS S79254 229-bp DNA BCT 30-NOV-1995
DEFINITION emm-M protein [Streptococcus pyogenes, M1a, Genomic, 229 nt].
ACCESSION S79254
VERSION S79254.1 GI:1087104
KEYWORDS Streptococcus pyogenes M1a.
SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 229)
AUTHORS Penney,T.J., Martin,D.R., Williams,L.C., de Malmarche,S.A. and Bergquist,P.L.
TITLE A single emm gene-specific oligonucleotide probe does not recognise all members of the Streptococcus pyogenes M type 1
JOURNAL FEMS Microbiol. Lett. 130 (2-3), 145-149 (1995)
MEDLINE 95377598
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 169934] from the original journal article.
This sequence comes from Fig. 1.
FEATURES
source
location/Qualifiers
1..229
/organism="Streptococcus pyogenes"
/db_xref="taxon:1314"
2..229
/partial
/gene="emm"
/note="M protein"
2..229
/partial
/gene="emm"
/note="This sequence comes from Fig. 2"
/codon_start=1
/product="M protein"
/protein_id="AAB35162.1"
/db_xref="GI:1087105"
/translation="SVAVRLTVLCAGFANOTEVKANGDNPREVIEDLAANNPAIQNI
RLRHENKDLKARLENMEVAGRDFKRAEELEK"
BASE COUNT 84 a 32 c 59 g 54 t
ORIGIN

Query Match 13.0%; Score 151; DB 3; Length 229;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 598 aacggtgatgtaacctaggaagttagaagatttcgcagcaaaacaaatcccgcaata 657
|||||
Db 65 AACGGTGATGTTAATCCTAGGGAAGTTATAGAAGATCTTGACGACAAACATCCCGCAATA 124

Qy 658 caaaataacgtttacgtcacgaaacaaagacctaaagcgagattagagaatgcaatg 717
|||||
Db 125 CAAAATATACGTTTACGTGCAGCAAAACAGGACTTAAAGCGAGATTAGAGAAATGCAATG 184

Qy 718 gaagttgcaggaagagattttaagagagctg 748
|||||
Db 185 GAAGTTGCAGGAGAGATTTTAAGAGAGCTG 215

RESULT 6
SPU11940
LOCUS SPU11940 237 bp DNA BCT 16-MAY-1996
DEFINITION Streptococcus pyogenes M type 1 (emm1) gene, partial cds.
ACCESSION U11940
VERSION U11940.1 GI:533557
KEYWORDS

SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 300)
AUTHORS Musser,J.M., Kapur,V., Szeto,J., Pan,X., Swanson,D.S. and Martin,D.R.
TITLE Genetic diversity and relationships among Streptococcus pyogenes

```

```

SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 237)
AUTHORS Whatmore,A.M.
TITLE Sequence Analysis of the Emm-Like Gene Family of Streptococcus pyogenes
JOURNAL Thesis (1993) Microbiology, University of Newcastle Upon Tyne
AUTHORS Whatmore,A.M., Kapur,V., Sullivan,D.J., Musser,J.M. and Kehoe,M.A.
TITLE Non-congruent relationships between variation in emm gene sequences and the population genetic structure of group A streptococci
JOURNAL Mol. Microbiol. 14 (4), 619-631 (1994)
MEDLINE 95198537
REFERENCE 3 (bases 1 to 237)
AUTHORS Whatmore,A.M.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1994) Embl Data Library By: A. M. Whatmore, Microbiology, University of Newcastle Upon Tyne, Framlington Place, Newcastle Upon Tyne, UK, NE2 4HH
FEATURES
source
location/Qualifiers
1..237
/organism="Streptococcus pyogenes"
/strain="M type 1"
/isolate="NCTC8198"
/db_xref="taxon:1314"
1..237
/gene="emm1"
<1..>237
/gene="emm1"
/codon_start=1
/transl_table=11
/protein_id="AAA99556.1"
/db_xref="GI:533558"
/translation="KTGTASVAVALTVLCAGFANOTEVKANGDNPREVIEDLAANNP
AIQNIRLRHENKDLKARLENMEVAGRDFKRAEEL"
BASE COUNT 86 a 35 c 62 g 54 t
ORIGIN

Query Match 13.0%; Score 151; DB 3; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 598 aacggtgatgtaacctaggaagttagaagatttcgcagcaaaacaaatcccgcaata 657
|||||
Db 79 AACGGTGATGTTAATCCTAGGGAAGTTATAGAAGATCTTGACGACAAACATCCCGCAATA 138

Qy 658 caaaataacgtttacgtcacgaaacaaagacctaaagcgagattagagaatgcaatg 717
|||||
Db 139 CAAAATATACGTTTACGTGCAGCAAAACAGGACTTAAAGCGAGATTAGAGAAATGCAATG 198

Qy 718 gaagttgcaggaagagattttaagagagctg 748
|||||
Db 199 GAAGTTGCAGGAGAGATTTTAAGAGAGCTG 229

RESULT 7
SPU20094
LOCUS SPU20094 300 bp DNA BCT 05-JAN-1996
DEFINITION Streptococcus pyogenes M1.0 protein gene, partial cds.
ACCESSION U20094
VERSION U20094.1 GI:643545
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE 1 (bases 1 to 300)
AUTHORS Musser,J.M., Kapur,V., Szeto,J., Pan,X., Swanson,D.S. and Martin,D.R.
TITLE Genetic diversity and relationships among Streptococcus pyogenes

```

strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease
Infect. Immun. 63 (3), 994-1003 (1995)
951/2752
REFERENCE 2 (bases 1 to 300)
AUTHORS Swanson, D.S.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1995) Douglas S. Swanson, Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COLLEGE OF MEDICINE, ONE BAYLOR PLAZA, HOUSTON, TX 77030, USA
LOCATION/QUALIFIERS
FEATURES
1. .300
/organism="Streptococcus pyogenes"
/isolate="MGAS 2116"
/db_xref="taxon:1314"
<1..>300
/codon_start=1
/transl_table=11
/product="M1.0 protein"
/protein_id="AA085107.1"
/db_xref="GI:643546"
/translation="TVLGGAFANQTEVKANGDNPREVIEDLAANNPAIONIRLHEN KDLKRLNEMEVAGDFKRAEELEKAKQALEQDKLDLETKLKLQDYDLAKESI"
BASE COUNT 126 a 45 c 67 g 62 t
ORIGIN
Query Match 13.0%; Score 151; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 598 aacggtgagtggaatcctaggaagattatagaagatcttgcagcaaaacatcccgcaata 657
Db 46 AACGGTGATGGTAACTCCTAGGCAAGTTATAGAAGATCTTGCAAGCAAAACATCCGCAATA 105
Qy 658 caaataatcgtttacgtcagcaaaacagagcacttaagcgaattagagaatgcaatg 717
Db 106 CAAAATATACGTTTACGTCACCAACAAAGGACTTAAAGCGAGATTAGAGATGCAATG 165
Qy 718 gaagttgcaggaagagattttaagagagctg 748
Db 166 GAAGTTGCAGGAAGAGATTTTAAGAGAGCTG 196
RESULT 8
AF288752 817 bp DNA BCT 02-AUG-2000
LOCUS Streptococcus pyogenes isolate 3148-00 M protein (emm) gene,
partial cds.
DEFINITION
ACCESSION AF288752
VERSION AF288752.1 GI:9652381
KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 817)
AUTHORS Beall, B.W.
TITLE 5' emm sequence emm3.2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 817)
AUTHORS Beall, B.W.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2000) Respiratory Diseases Branch, Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, Atlanta, GA 30333, USA
LOCATION/QUALIFIERS
FEATURES
1. .817
/organism="Streptococcus pyogenes"
/isolate="3148-00"
/db_xref="taxon:1314"
/note="blood isolate recovered from patient in New York population based surveillance; f non typeable; opacity factor negative"

group: A"
<1..>817
/gene="emm"
<1..>817
/gene="emm"
/note="emm3.2"
/codon_start=1
/transl_table=11
/product="M protein"
/protein_id="AAF91496.1"
/db_xref="GI:9652382"
/translation="TRNSVAVALTVLGTGLVAGQTVKADARSVDARSVNGEPRHVKLKN EIEILLQVLTQITKHSNYSQYNAQAGRLDLRQAEYLKGLNDAERLLGELNGEDV KKLGVAFPEKDDLEKKEKEDYQDLKDFDLAKOGYVLSDRHQOELE EKEKRVTEATKVGQISBELETVKQKVEKMODLTEKQNRYSQLEQELATTKQNAKED PELAAANAADKQKLEAKTADLETKLKEAKEDFELAALGHQHAHNEYQAKLAEK"
BASE COUNT 342 a 105 c 182 g 188 t
ORIGIN
Query Match 13.0%; Score 151; DB 2; Length 817;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 753 cttgttagatcagggttacacaattatataactaataataatagtaattaccacaataataa 812
Db 144 CTTGTTAGATCAGGTTACACAATTATATATACTAAACATAATAGTAATTACCAACATATAA 203
Qy 813 tgcacaagctggcagactgacctgagacaaaagcgtgaatctctaaaagcgttaataa 872
Db 204 TGCACAGCTGGCAGACTTGACCTGAGACAAAAGGCTGATATCTAAAGGCTTTAATGA 263
Qy 873 ttgggctgagagcgtgttacaagagtgtaaat 903
Db 264 TTGGCTGAGAGGCTGTGTACAAGAGTAAAT 294
RESULT 9
AR079002 1332 bp DNA PAT 31-AUG-2000
LOCUS Sequence 5 from patent US 5965390.
DEFINITION
ACCESSION AR079002
VERSION AR079002.1 GI:10005748
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1332)
AUTHORS Bjorck, L. and Sjobring, U.
TITLE Protein L and hybrid proteins thereof
JOURNAL Patent: US 5965390-A 5 12-OCT-1999;
FEATURES
source Location/Qualifiers
1. .1332
/organism="unknown"
BASE COUNT 568 a 244 c 281 g 239 t
ORIGIN
Query Match 13.0%; Score 151; DB 9; Length 1332;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 598 aacggtgatgtaacctcagggaagttatagaagatcttgcagcaaaacatcccgcaata 657
Db 1 AACGGTGATGGTAACTCCTAGGCAAGTTATAGAAGATCTTGCAAGCAAAACATCCGCAATA 60
Qy 658 caaataatcgtttacgtcagcaaaacagagcacttaagcgaattagagaatgcaatg 717
Db 61 CAAAATATACGTTTACGTCACCAACAAAGGACTTAAAGCGAGATTAGAGATGCAATG 120
Qy 718 gaagttgcaggaagagattttaagagagctg 748
Db 121 GAAGTTGCAGGAAGAGATTTTAAGAGAGCTG 151

Qy	718	gaagtgcagagaagattttaacagagctg	748
Dd	274	GAAATTGCAGGAAGAGATTTTAAGAGAGCTG	304
RESULT 11			
SPU40231			
LOCUS	SPU40231	1620 bp	DNA
DEFINITION	Streptococcus pyogenes M3 protein (emm3.1) gene, BCT		15-APR-1996
ACCESSION	U40231		complete cds.
VERSION	U40231.1	GI:1263022	
KEYWORDS			
SOURCE	Streptococcus pyogenes strain=Type 3.		
ORGANISM	Streptococcus pyogenes Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcus;		Streptococcaceae;
REFERENCE	1 (bases 1 to 1620)		
AUTHORS	Dale,J.B., Simmons,M., Chiang,E. and Chiang,E.		
TITLE	Octavalent group A streptococcal M protein vaccine		
JOURNAL	Vaccine (1996) In press		
REFERENCE	2 (bases 1 to 1620)		
AUTHORS	Dale,J.B.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-NOV-1995) James B. Dale, Medicine, VA Medical Center, 1030 Jefferson Ave., Memphis, TN 38104, USA		
FEATURES	Location/Qualifiers		
source	1..1620		
gene	/organism="Streptococcus pyogenes" /strain="Type 3" /db_xref="taxon:1314"		
CDS	1..1620 /gene="emm3.1" 1..1620 /gene="emm3.1" /codon_start=1 /transl_table=11 /product="M3 protein" /protein_id="AAA96960.1" /db_xref="GI:1263023"		
BASE COUNT	697 a 269 c 335 g 319 t		
ORIGIN	/translation="MAKNNTNRHSYLRKLTGTASVAVALTVLGTGLVAGQTVKADAR SWNEPPIRVKLKNETJENLLDQVTQLYTKHNSNYOYNQAAGRDLDRKAEYLKGLND GYVLRLQELNGEDVKVKGAVKDDLEKEVKELKEIDKKEKYODLDKDFDLAK GVLSDKHQOELEEKVKVTATAKVGOI SELETVKOKVESTMODLTEKQNRVSOL EQLATT KONAKEDFELALANNAKQLEAKIADLTETLKEAKEOFELAALGHQHAH NSYQAKLAKSDQIKOLESQKQLDRAKRGTDHPAIVROAKKATEAEILNNLKAEIAH VTEQKLIQASRKRTARDLEAVRQAQAEALKQLEQNRI SEASKRGLRRDLDAER EAKQVEKALEANSKLALEKLKEESKLTEKEKELQAKLEAKAKLKEQLAK QAEELAKLAGKASDISIPDTKPGNKAVPGKQPAQGTPKPNONKAPMKETKKPLPST GETANPFPTAAALTVMATGAVAVVRKKEEN"		
Query Match	13.0%;	Score 151;	DB 3; Length 1620;
Best Local Similarity	100.0%;	Pred. No. 1.2e-69;	
Matches 151;	Conservative	0; Mismatches	0; Indels
			0; Gaps
0;			
QY	753	cttgtagatcagggttacacaattataactaacacaataatagtaattaccacaataaa	812
Dd	186	CTTGTTAGTACAGTTACACATTATATACTAACACAATAGTAATTACCACAATAATA	245
QY	813	tgcacaagctggcagacttgacctgagacaaaaggctggaatatctaaaaagcccttaatga	872
Dd	246	TGCACAAGCTGGCAGACTTGACCTGAGACAAAAAGCGTGAATATCTAAAGCCCTTAATGA	305
QY	873	ttggcgctgagaggctgttacaagagtgtaaatt	903
Dd	306	TTGGCGCTGAGAGGCTGTTCACCAAGAGTTAAAT	336
RESULT 12			
SPPEMM1			

LOCUS SPMM1 1708 bp DNA BCT 14-MAR-1996
DEFINITION S.pyogenes emm1 gene for type 1 group A M protein.
ACCESSION X07933 X07860
VERSION X07933.1 GI:47358
KEYWORDS emm1 gene; M protein.
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 1708)
AUTHORS Hannes-Fritz,E.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-1988) Hannes-Fritz E., University of Minnesota,
Department of Microbiology, Box 196, Mayo Building, Minneapolis, MN
55455
REFERENCE 2 (bases 1 to 1708)
AUTHORS Haanes-Fritz,E., Kraus,W., Burdett,V., Dale,J.B., Beachey,E.H. and
Cleary,P.
TITLE Comparison of the leader sequences of four group A streptococcal M
protein genes
JOURNAL Nucleic Acids Res. 16 (10), 4667-4677 (1988)
MEDLINE 88247768
FEATURES Location/Qualifiers
source
1..1708
/organism="Streptococcus pyogenes"
/strain="CS130"
/db_xref="taxon:1314"
/clones="pbf82"
921..963
/note="pot.promoter region"
1015..1019
1028..1708
/codon_start=1
/transl_table=11
/product="M protein precursor"
/protein_id="CAA30757.1"
/db_xref="GI:47359"
/db_xref="SPTREMBL:Q10371"
/translation="MAKNNTNRHSLKRLKTGTSVAVALTVLGAGFANTEYKANGD
GNPREVIEDLAANNPAIQNIIRHENDKRLKARLENAMVAGRODFKRAEELEKAKQALE
DQRDLTKLEQDYDLAKESTSWDRQLEKELEKKEALELAIQASRDYHRATA
LEKELEKKKALELAIQASQDYNRANVLEKELETTITREQEINRLNLGNKLELDQLS
SEKQLTIR"
sig_peptide 1028..1150
mat_peptide 1151..1708
repeat_region 1427..1525
repeat_region 1522..1609
/rpt_type=DIRECT
/rpt_type=DIRECT
BASE COUNT 640 a 310 c 331 g 427 t
ORIGIN

Query Match 13.0%; Score 151; DB 3; Length 1708;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 598 aacgggtgatgtaactcctaggaagttagagaagatcttgagcaaaccaatcccgcaata 657
|||||
DB 1151 AACGGTGATGGTAATCTAGGAGAGTTATAGAGATCTTGCGACCAACAATCCCGCAATA 1210
|||||
QY 658 caaaatatacgtttacgtcacgaacaacaggacttaaaagcgagattagagaatgcaatg 717
|||||
DB 1211 CAAAATATACGTTTAGGTACGCACGAAACAGGACTTAAAGCGAGATTAGAGAATGCAATG 1270
|||||
QY 718 gaagttgcggaagaagatttaagaagactg 748
|||||
DB 1271 GAAGTTGCAGGAGAGAGATTTAAGAGAGCTG 1301

RESULT 13
SPU20097

LOCUS SPU20097 300 bp DNA BCT 05-JAN-1996
DEFINITION Streptococcus pyogenes M1.3 protein gene, partial cds.
ACCESSION U20097
VERSION U20097.1 GI:643551
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 300)
AUTHORS Musser,J.M., Kapur,V., Szeto,J., Pan,X., Swanson,D.S. and
Martin,D.R.
TITLE Genetic diversity and relationships among Streptococcus pyogenes
strains expressing serotype M1 protein: recent intercontinental
spread of a subclone causing episodes of invasive disease
Infect. Immun. 63 (3), 994-1003 (1995)
JOURNAL 95172752
MEDLINE 2 (bases 1 to 300)
REFERENCE Swanson,D.S.
AUTHORS Direct Submission
TITLE Submitted (19-JAN-1995) Douglas S. Swanson, Pathology, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
JOURNAL
FEATURES Location/Qualifiers
source
1..300
/organism="Streptococcus pyogenes"
/isolate="MGAS 249"
/db_xref="taxon:1314"
<1..>300
/codon_start=1
/transl_table=11
/product="M1.3 protein"
/protein_id="AA85110.1"
/db_xref="GI:643552"
/translation="TVLGAGFANQTEYKANGDSPREVIEDLAANNPAIQNIIRHEN
KDLKARLENAMVAGRODFKRAEELEKAKQALEQDKLTKLEQDYDLAKEST"
BASE COUNT 125 a 45 c 68 g 62 t
ORIGIN

Query Match 11.8%; Score 137; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 3.9e-62;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 612 tcttagggaggttatagaagatcttgcagcaaacatccgcaatacaataatatacgttt 671
|||||
DB 60 TCTTAGGGAAGTTATAGAGATCTTGCGACCAACAATCCCGCAATACAAATATACGTTT 119
|||||
QY 672 acgtcacgaaaaaacgaacttaaaagcgagattagagaatgcaatggaattgcaggaag 731
|||||
DB 120 ACGTCACGAAAACAGGACTTAAAGCGAGATTAGAGAAATGCAATGGAAGTTGCAGGAAG 179
|||||
QY 732 agattttaagagagctg 748
|||||
DB 180 AGATTTTAAGAGAGCTG 196

RESULT 14
AF288753
LOCUS AF288753 694 bp DNA BCT 02-AUG-2000
DEFINITION Streptococcus pyogenes isolate 7203-99 M protein (emm) gene,
partial cds.
ACCESSION AF288753
VERSION AF288753.1 GI:9652383
KEYWORDS Streptococcus pyogenes.
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 694)
AUTHORS Beall,B.W.
TITLE 5' emm sequence emm3.3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 694)


```
AUTHORS Beall,B.W.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2000) Respiratory Diseases Branch, Centers for
Disease Control and Prevention, 1600 Clifton Rd. NE, Atlanta, GA
30333, USA
FEATURES
Source Location/Qualifiers
1..694
/organism="Streptococcus pyogenes"
/isolate="7203-99"
/db_xref="taxon:1314"
/note="type 3; opacity factor negative; blood isolate"
group: A"
<1..>694
/gene="emm"
<1..>694
/gene="emm"
/note="emm3.3"
/codon_start=1
/transl_table=11
/product="M protein"
/protein_id="AAF91497.1"
/db_xref="GI:9652384"
/translation="TASVAVALTVLGTLVAGQTVKADARSVNGEFPRHVKLNEIEN
LLDOVTQLYTKHIVTKHNSYQYNAGRLDLRQAEYLKGLNDWAEERLQELNGED
VKVLGKVAEFKEDLEKEVKELEKIDKKEKEYODLDRKFDLAKQGVLSDKRHQDEL
EEKKKVTAEAKVQISEELETVKQVESTMQDLTEKQNRVSOLEQELATTQNAKE
DFELAAANAADK"
BASE COUNT 289 a 87 c 152 g 166 t
ORIGIN

Query Match 11.1%; Score 128; DB 2; Length 694;
Best Local Similarity 100.0%; Pred. No. 2.6e-57;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 tataactaaacataatagtaattaccacaaataataatgcacagctgcagacttgacc 835
Db 170 TATATACTAAACATAATAGTATTACCAACATATATATGACACAGCTGGCAGACTTGACC 229

QY 836 ttagacaaaggctgaatctctaaaggccttaatgattggcgtgagaggctgttacaag 895
Db 230 TGAGACAAAAGCGTGAATATCTATAAAGCGCTTAATGATTGGCTGAGAGGCTGTACAAAG 289

QY 896 agttaaat 903
Db 290 AGTTAAAT 297

RESULT 15
SPU20099
LOCUS SPU20099 300 bp DNA BCT 05-JAN-1996
DEFINITION Streptococcus pyogenes M1.5 protein gene, partial cds.
ACCESSION U20099
VERSION U20099.1 GI:643555
KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes.
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
1 (bases 1 to 300)
Musser,J.M., Kapur,V., Szeto,J., Pan,X., Swanson,D.S. and
Martin,D.R.
Genetic diversity and relationships among Streptococcus pyogenes
strains expressing serotype M1 protein: recent intercontinental
spread of a subclone causing episodes of invasive disease
Infect. Immun. 63 (3), 994-1003 (1995)
95172752
2 (bases 1 to 300)
Swanson,D.S.
Direct Submission
AUTHORS Swanson,D.S.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-1995) Douglas S. Swanson, Pathology, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
FEATURES
Location/Qualifiers
```

```
source 1..300
/organism="Streptococcus pyogenes"
/isolate="MGAS 1666"
/db_xref="taxon:1314"
<1..>300
/codon_start=1
/transl_table=11
/product="M1.5 protein"
/protein_id="AAA85112.1"
/db_xref="GI:643556"
/translation="TVLGAGFANQTEVKANGDNPREVIEGLAANPAIONIRLRHEN
KDLKARLENAMEVAGRDFKRAELEKAKALEDKRKLETKELQOQDYDLAKEST"
BASE COUNT 125 a 45 c 68 g 62 t
ORIGIN

Query Match 10.0%; Score 116; DB 3; Length 300;
Best Local Similarity 100.0%; Pred. No. 7e-51;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 633 tcttgacgacaaacaatcccgaatacacaaaataatacgtttacgtcacgaaacaaaggactt 692
Db 81 TCTTGCAGCAACAAATCCCGCAATACAAAATATACGTTTACGTACCAAGAAACAAAGGACTT 140

QY 693 aaaagcgagattagagaaatgcaatggaagttgcagggaagagattttaagagagactg 748
Db 141 AAAAGCGAGATTAGAGAATGCAATGGAAGTTGCAGGAAGAGAGATTTTAAGAGAGACTG 196
```

Search completed: May 13, 2001, 06:13:28
Job time: 4538 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:59:10 ; Search time 121.03 Seconds
(without alignments)
5585.547 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcgatggtcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 678276 seqs, 291890651 residues
Word size : 10

Total number of hits satisfying chosen parameters: 214117

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N_Geneseq_0401:*
1: /SID22/gcgdata/geneseq/NA1980.DAT:*
2: /SID22/gcgdata/geneseq/NA1981.DAT:*
3: /SID22/gcgdata/geneseq/NA1982.DAT:*
4: /SID22/gcgdata/geneseq/NA1983.DAT:*
5: /SID22/gcgdata/geneseq/NA1984.DAT:*
6: /SID22/gcgdata/geneseq/NA1985.DAT:*
7: /SID22/gcgdata/geneseq/NA1986.DAT:*
8: /SID22/gcgdata/geneseq/NA1987.DAT:*
9: /SID22/gcgdata/geneseq/NA1988.DAT:*
10: /SID22/gcgdata/geneseq/NA1989.DAT:*
11: /SID22/gcgdata/geneseq/NA1990.DAT:*
12: /SID22/gcgdata/geneseq/NA1991.DAT:*
13: /SID22/gcgdata/geneseq/NA1992.DAT:*
14: /SID22/gcgdata/geneseq/NA1993.DAT:*
15: /SID22/gcgdata/geneseq/NA1994.DAT:*
16: /SID22/gcgdata/geneseq/NA1995.DAT:*
17: /SID22/gcgdata/geneseq/NA1996.DAT:*
18: /SID22/gcgdata/geneseq/NA1997.DAT:*
19: /SID22/gcgdata/geneseq/NA1998.DAT:*
20: /SID22/gcgdata/geneseq/NA1999.DAT:*
21: /SID22/gcgdata/geneseq/NA2000.DAT:*
22: /SID22/gcgdata/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1158	100.0	1158	20	X33103 Streptococcal hexa
2	402	34.7	741	21	A57894 S. pyogenes hybrid
3	351	30.3	741	15	Q45215 Recombinant M24-M5
4	240	20.7	861	15	Q45214 Recombinant M24-M5
5	240	20.7	861	21	A57893 Recombinant M24-M5
6	217	18.7	522	15	Q45220 S. pyogenes hybrid
7	217	18.7	522	21	A57899 Recombinant M24-M5
8	217	18.7	741	15	Q45219 S. pyogenes hybrid
9	217	18.7	741	21	A57898 Recombinant M19-M6
10	217	18.7	822	15	Q45216 S. pyogenes hybrid
11	217	18.7	822	21	A57895 Recombinant M24-M5
					S. pyogenes hybrid

12	217	18.7	1029	15	Q45221
13	217	18.7	1029	21	A57900
14	105	9.1	1617	18	T49317
15	63	5.4	301	6	N50341
16	53	4.6	561	15	Q45217
17	53	4.6	561	21	A57896
18	52	4.5	204	15	Q45222
19	52	4.5	204	21	A57901
20	52	4.5	408	15	Q45223
21	52	4.5	408	21	A57902
22	52	4.5	918	15	Q45218
23	52	4.5	918	21	A57897
24	48	4.1	765	15	Q45160
25	48	4.1	855	15	Q45161
26	38	3.3	39	21	A57905
27	36	3.1	417	15	Q45159
28	34	2.9	47	21	A57907
29	32	2.8	54	21	A57917
30	27	2.3	33	20	X33102
31	25	2.2	30	15	Q45210
32	25	2.2	30	20	X30302
33	25	2.2	30	21	A57913
34	25	2.2	36	20	X30298
35	24	2.1	30	15	Q45208
36	24	2.1	30	15	Q45209
37	24	2.1	30	15	Q45212
38	24	2.1	30	15	Q45213
39	24	2.1	30	20	X33099
40	24	2.1	30	20	X33100
41	24	2.1	30	20	X33101
42	24	2.1	30	20	X30299
43	24	2.1	30	20	X30300
44	24	2.1	30	20	X30301
45	24	2.1	30	20	X30303
46	24	2.1	30	20	X30304
47	24	2.1	30	20	X30305
48	24	2.1	30	20	X33098
49	24	2.1	30	21	A57911
50	24	2.1	30	21	A57918
51	24	2.1	30	21	A57915
52	24	2.1	30	21	A57916
53	22	1.9	29	15	Q45206
54	22	1.9	29	21	A57903
55	22	1.9	29	21	A57909
56	22	1.9	32	20	X03263
57	19	1.6	30	21	A57918
58	19	1.6	390	16	T23376
59	19	1.6	2367	21	C42916
60	18	1.6	784	18	V74848
61	18	1.6	1260	21	C64427
62	18	1.6	1392	18	T72874
63	18	1.6	5836	20	X86366
64	18	1.6	10336	20	X34652
65	17	1.5	623	21	C35048
66	17	1.5	719	21	F12228
67	17	1.5	994	21	C34777
68	17	1.5	995	21	C35114
69	17	1.5	1090	20	X13612
70	17	1.5	1109	21	C54025
71	17	1.5	1120	21	C59916
72	17	1.5	1440	21	Z37035
73	17	1.5	1866	19	V53502
74	17	1.5	2858	18	T62650
75	17	1.5	2858	18	T64391
76	17	1.5	2858	18	T59328
77	17	1.5	3096	8	N71063
78	17	1.5	3929	21	F21072
79	17	1.5	3929	21	A34950
80	17	1.5	4668	19	V60918
81	17	1.5	5077	16	Q97812
82	17	1.5	5480	21	X13069
83	17	1.5	5912	21	C75806
84	17	1.5	9834	18	V74348

Tetavalent-C repe
S. pyogenes hybrid
Type-6 M-protein e
Sequence of a port
Recombinant M24-M5
S. pyogenes hybrid
M24-M5-M6-M19 hybr
S. pyogenes hybrid
M24-M5-M6-M19-M3-M
S. pyogenes octava
Recombinant M24-M5
S. pyogenes hybrid
Sequence of a frag
Sequence of a fragme
Streptococcus pyog
Sequence of LT-B-M
Streptococcus pyog
Streptococcus pyog
Streptococcal hexa
M6 top strand prim
Streptococcal hexa
Streptococcus pyog
Streptococcal hexa
M5 bottom strand p
M19 top strand pri
M19 bottom strand
Streptococcal hexa
Streptococcal hexa
Streptococcal hexa
Streptococcal hexa
Streptococcal hexa
Streptococcal hexa
Streptococcal hexa
Streptococcus pyog
Streptococcus pyog
Streptococcus pyog
M24 top strand pri
Streptococcus pyog
Streptococcus pyog
PCR primer used to
Human gene signatu
Arabidopsis thalia
Staphylococcus aur
Staphylococcus aur
DNA encoding glycy
DNA encoding the h
Wheat starch solub
Arabidopsis thalia
Aspergillus oryzae
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human secreted pro
DNA of the agAB ge
DNA encoding a sta
DNA encoding OCHI
Pichia pastoris al
OCHI gene. Pichia
Genomic sequence e
Human low adenosin
Human adenosine re
Angiogenin gene.
Sequence encoding
Enterococcus faeca
Human ORFX ORF1361
Staphylococcus aur

c 961 14 1.2 229 15 Q77397 Human genome fragm
c 962 14 1.2 229 21 A45299 Human secreted exp
c 963 14 1.2 233 14 Q37282 SSP-seg34 from clo
c 964 14 1.2 240 21 C10392 Human secreted pro
c 965 14 1.2 242 21 C02545 Human secreted pro
c 966 14 1.2 243 21 A15442 DNA sequence whose
c 967 14 1.2 245 21 C09244 Human secreted pro
c 968 14 1.2 250 21 C20216 Human secreted pro
c 969 14 1.2 251 18 V78866 Staphylococcus aur
c 970 14 1.2 251 20 X84221 DNA encoding human
c 971 14 1.2 251 20 X84222 EST clone FM32. H
c 972 14 1.2 251 21 C79567 Known human breast
c 973 14 1.2 251 21 C79450 Known human breast
c 974 14 1.2 251 21 C79451 Human secreted pro
c 975 14 1.2 251 21 C16806 Human gene signatu
c 976 14 1.2 252 16 T25775 Human gene signatu
c 977 14 1.2 252 16 T22061 Human brain Expres
c 978 14 1.2 252 16 Q60621 Nucleotide sequenc
c 979 14 1.2 257 20 V90679 Human secreted pro
c 980 14 1.2 258 21 C26740 Pinus radiata C4H
c 981 14 1.2 259 21 A68064 Human secreted pro
c 982 14 1.2 259 21 C07909 Human brain Expres
c 983 14 1.2 260 21 Q60636 Human secreted pro
c 984 14 1.2 260 21 C25246 Senescence-related
c 985 14 1.2 261 17 T28145 EST clone BG374
c 986 14 1.2 275 20 V86815 Human secreted pro
c 987 14 1.2 276 21 C17731 Human secreted pro
c 988 14 1.2 283 21 C29120 Human secreted pro
c 989 14 1.2 285 20 X40546 Human secreted pro
c 990 14 1.2 285 21 C23147 Human secreted pro
c 991 14 1.2 286 13 Q29025 Low frequency repe
c 992 14 1.2 286 21 C14492 Human secreted pro
c 993 14 1.2 287 21 C23517 Human secreted pro
c 994 14 1.2 297 18 T91735 Mouse SMLIM/CRP2 g
c 995 14 1.2 297 21 C64711 SMLIM/CRP2 related
c 996 14 1.2 300 20 Z14998 Human gene express
c 997 14 1.2 300 20 Z14520 Human gene express
c 998 14 1.2 300 20 Z13859 Human gene express
c 999 14 1.2 300 20 Z13732 Human gene express
c1000 14 1.2 300 20 Z13485 Human gene express

ALIGNMENTS

RESULT 1
ID X33103 standard; DNA; 1158 BP.
XX
AC X33103;
XX
DT 23-JUN-1999 (first entry)
XX
DE Streptococcal hexavalent M protein vaccine encoding DNA.
XX
KW Streptococcal; hexavalent emm gene; multivalent; vaccine; Group A;
KW immunogenic; immune response; pathogen; fusion protein; ss.
OS Streptococcus sp.
XX
PN W09913084-A1.
XX
PD 18-MAR-1999.
XX
PF 14-SEP-1998; 98WO-US19100.
XX
PR 12-SEP-1997; 97US-0058635.
XX
FA (IDVA-) ID VACCINE.
XX
PI Dale JB;
XX
DR WPI; 1999-215066/18.

DR P-PSDB; Y04368.
XX Immunogenic fusion protein derived from group A streptococci
PT
XX
PS Example 1; Fig 7; 50pp; English.
XX
CC The present invention describes an immunogenic fusion polypeptide (I)
CC that stimulates an immune response against group A streptococci (GAS).
CC (I) comprises: (a) at least 2 peptides (Ia) from GAS, at least 10 amino
CC acids long and (able to induce a specific immune response; and (b) a
CC peptide (Ib), C-terminal to (Ia) that protects the immunogenicity of the
CC component described in (a) but is not essential for stimulation of the
CC immune response. Vaccines containing (I) are used to protect against GAS
CC infections, specifically those caused by Streptococcus pyogenes, e.g.
CC pharyngitis, pyoderma, toxic shock syndrome, deep tissue invasion, sepsis
CC and acute rheumatic fever. (I) have improved immunogenicity and do not
CC generate antibodies that are cross-reactive with human tissues. The
CC present sequence encodes a hexavalent M protein vaccine.
XX
SQ Sequence 1158 BP; 474 A; 165 C; 255 G; 264 T; 0 other;

Query Match 100.0%; Score 1158; DB 20; Length 1158;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gcatgcatggtcgagctaggtctcagacagatctctggaaaaatacaagaactgct 60
Db 1 gcatgcatggtcgagctaggtctcagacagatctctggaaaaatacaagaactgct 60
Qy 61 gacaagtttgagatagaaaaacaatacattaaacttaagaatagtagtaagttta 120
Db 61 gacaagtttgagatagaaaaacaatacattaaacttaagaatagtagtaagttta 120
Qy 121 aataaagcgttaaaagatcataatgatgagtaagtaagtagtgaatgtaagag 180
Db 121 aataaagcgttaaaagatcataatgatgagtaagtaagtagtgaatgtaagag 180
Qy 181 aaactagtggtatccgcgtgactaggggtacataataatgacccgcgaagagaa 240
Db 181 aaactagtggtatccgcgtgactaggggtacataataatgacccgcgaagagaa 240
Qy 241 gctctgacaagtatgagctagaaaaacacatgacttaaaactaagaatgaaggt 300
Db 241 gctctgacaagtatgagctagaaaaacacatgacttaaaactaagaatgaaggt 300
Qy 301 actgagaatgaaggtttaaaactgagaatgaaggttaaaactgagaatgaaggt 360
Db 301 actgagaatgaaggtttaaaactgagaatgaaggttaaaactgagaatgaaggt 360
Qy 361 aaactagtggtcagacagagtggttctcctaggggacggttagaaaaacccgg 420
Db 361 aaactagtggtcagacagagtggttctcctaggggacggttagaaaaacccgg 420
Qy 421 gaactcttaacaagtatgacgtagagaactctatgtttacaagtaataatgaca 480
Db 421 gaactcttaacaagtatgacgtagagaactctatgtttacaagtaataatgaca 480
Qy 481 ccatggagagtcgtttactactagcaccagacagatgaagctaaataattattg 540
Db 481 ccatggagagtcgtttactactagcaccagacagatgaagctaaataattattg 540
Qy 541 gatcttgacgcaaaagacatgaattacaacaagaatgagaagtattctctcaga 600
Db 541 gatcttgacgcaaaagacatgaattacaacaagaatgagaagtattctctcaga 600
Qy 601 ggtatggttaattccttaggggaagttatagaagatcttgacgaacaacatcc 660
Db 601 ggtatggttaattccttaggggaagttatagaagatcttgacgaacaacatcc 660
Qy 661 aatatacgtttacgtcacgaaacaagagacttaaaagcgagattagagaatg 720
Db 661 aatatacgtttacgtcacgaaacaagagacttaaaagcgagattagagaatg 720

QY 721 gttcagagaagattttaagagagcgtgtaccttgttagatcaggttacacaattatat 780
Db |||||||
QY 721 gttcagagaagattttaagagagcgtgtaccttgttagatcaggttacacaattatat 780
Db |||||||
QY 781 actaaacataatagtaattaccaacaataataatgacaaagctggcagacttgacctgaga 840
Db |||||||
QY 781 actaaacataatagtaattaccaacaataataatgacaaagctggcagacttgacctgaga 840
Db |||||||
QY 841 caaaagcgtgaatatctaaagccttaattgattggctgagagcgtgttacaagagtta 900
Db |||||||
QY 841 caaaagcgtgaatatctaaagccttaattgattggctgagagcgtgttacaagagtta 900
Db |||||||
QY 901 aatatcgatgtcgcactaggtctcagacagatactctggaaaaagtacaaagaacgtgct 960
Db |||||||
QY 901 aatatcgatgtcgcactaggtctcagacagatactctggaaaaagtacaaagaacgtgct 960
Db |||||||
QY 961 gacaagtttgatagaaaacataactgttaaaacttaagaatactgacttaagttttaatt 1020
Db |||||||
QY 961 gacaagtttgatagaaaacataactgttaaaacttaagaatactgacttaagttttaatt 1020
Db |||||||
QY 1021 aataaagcgttaaaagataataatgatgagtttaactgaagcgttgagtaagtctaaagag 1080
Db |||||||
QY 1021 aataaagcgttaaaagataataatgatgagtttaactgaagcgttgagtaagtctaaagag 1080
Db |||||||
QY 1081 aaactacgtataaataatgataatcactatctgaaaaagctagtaaaattcaagatttagag 1140
Db |||||||
QY 1081 aaactacgtataaataatgataatcactatctgaaaaagctagtaaaattcaagatttagag 1140
Db |||||||
QY 1141 gcacgtgaagtaaaagcctt 1158
Db |||||||
QY 1141 gcacgtgaagtaaaagcctt 1158
Db |||||||

RESULT 2
A57894
ID A57894 standard; DNA; 741 BP.
XX
AC A57894;
XX
DT 10-OCT-2000 (first entry)
XX
DE S. pyogenes hybrid M protein (M24-M5-M6-M19) DNA, SEQ ID NO:3.
XX
KW Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia; ds.
XX
OS Streptococcus pyogenes.
OS Synthetic.
XX
PN US6063386-A.
XX
PD 16-MAY-2000.
XX
PF 15-SEP-1997; 97US-0937271.
XX
PR 16-SEP-1992; 92US-0945954.
XX
PA (UYTE-) UNIV TENNESSEE RES CORP.
XX
PI Lederer JW, Dale JB;
XX
DR WPI: 2000-364475/31.
DR P-PSDB; B03114.
XX
PT New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX
PS Disclosure; Fig 4A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences A57893-A57902
CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
CC proteins generated in the disclosure of the invention.
SQ Sequence 741 BP; 311 A; 106 C; 164 G; 160 T; 0 other;

Query Match 34.7%; Score 402; DB 21; Length 741;
Best Local Similarity 100.0%; Pred. No. 4.4e-190;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 190 ggatccgcgtgactagggttacataataatgaccccaagacgaagaagctcttgac 249
Db |||||||
QY 340 ggatccgcgtgactagggttacataataatgaccccaagacgaagaagctcttgac 399
Db |||||||
QY 250 aagtatgactagaaaaccatgacttaaaactaagaatgaaggggttaaaactgagaat 309
Db |||||||
QY 400 aagtatgactagaaaaccatgacttaaaactaagaatgaaggggttaaaactgagaat 459
Db |||||||
QY 310 gaaggggttaaaactgagaatgaaggggttaaaactgagaatgaaggggttaaaactgaga 369
Db |||||||
QY 460 gaaggggttaaaactgagaatgaaggggttaaaactgagaatgaaggggttaaaactgaga 519
Db |||||||
QY 370 gtcgacagagtgttctctaggggacggtagaaaaccccgacagacgagaactctt 429
Db |||||||
QY 520 gtcgacagagtgttctctaggggacggtagaaaaccccgacagacgagaactctt 579
Db |||||||
QY 430 acaagtgatgactagagaactctatgttacagctaatataatgacaagtacatggaga 489
Db |||||||
QY 580 acaagtgatgactagagaactctatgttacagctaatataatgacaagtacatggaga 639
Db |||||||
QY 490 gtgcgttactaggcatacgcagaagaataagctaaaaaaattattgacgactctgac 549
Db |||||||
QY 640 gtgcgttactaggcatacgcagaagaataagctaaaaaaattattgacgactctgac 699
Db |||||||
QY 550 gcaaaagaacatgaattacaacaacagaatgagaagttatct 591
Db |||||||
QY 700 gcaaaagaacatgaattacaacaacagaatgagaagttatct 741
Db |||||||

RESULT 3
Q45215
ID Q45215 standard; DNA; 741 BP.
XX
AC Q45215;
XX
DT 02-NOV-1994 (first entry)
XX
DE Recombinant M24-M5-M6-M19 gene.
XX

KW	Primer: PCR: amplify; polymerase chain reaction; construct; hybrid;	
KW	M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;	
KW	tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;	
KW	restriction enzyme site; multivalent M protein; immunisation; group A;	
KW	streptococci; rheumatic fever; rheumatic heart disease; humoral;	
KW	antibody; heart tissue; antigen; serotype; mucosal; ss.	
XX	Synthetic.	
XX	WO9406421-A.	
PN	31-MAR-1994.	
PD	15-SEP-1993; 93WO-US08703.	
PF	16-SEP-1992; 92US-0945954.	
XX	(UYTE-) UNIV TENNESSEE RES CORP.	
XX	Dale JB, Lederer JW;	
XX	WPI: 1994-118122/14.	
DR	P-PSDB; R50993.	
DR	New immunogenic hybrid proteins derives from streptococcal M proteins	
PT	- induces opsonic antibodies, for protective immunisation against	
PT	against multiple group A streptococci serotypes	
XX	Disclosure; Fig 4; 67pp; English.	
XX	The sequences given in Q45214-23 encode hybrid M proteins which	
CC	contain the M24-M5-M6 and/or M19 subunits. These tetraivalent	
CC	proteins were constructed using fragments of the 5' regions of emm	
CC	genes that were amplified by PCR, ligated in tandem and expressed in	
CC	PKK223.3. The amplified regions pref. encode protective and not	
CC	tissue-cross-reactive epitopes, which can then be linked into one	
CC	protein molecule. The recombinant hybrid protein may contain 113	
CC	N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and	
CC	35 from M19. Each section is linked by 2 amino acids specified by	
CC	the respective restriction enzyme sites that were synthesised into	
CC	the primers used to specify the PCR product. Multivalent M proteins	
CC	such as this may be used for protective immunisation against group A	
CC	streptococci, which esp. cause rheumatic fever and rheumatic heart	
CC	disease. Humoral antibodies raised against these proteins do not	
CC	react with heart tissue antigens but are effective against many	
CC	different serotypes. The multivalent proteins may also include	
CC	sequences which induce mucosal antibodies and do not require coupling	
CC	to an immunogenic carrier.	
XX	Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;	
SQ		
QY	Query Match 30.38; Score 351; DB 15; Length 741;	
DB	Best Local Similarity 99.88; Pred. No. 9.4e-165;	
DB	Matches 401; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	190 gqatccgcgtgactaggggtacataataatgaccgcgcaagagcaaaaagagctcttgac 249	
DB		
DB	340 gqatccgcgtgactaggggtacataataatgaccgcgcaagagcaaaaagagctcttgac 399	
QY	250 aagtgatgactagaaaacacagcttaaaaactaagaatgaagggttaaaaactgagaat 309	
DB		
DB	400 aagtgatgactagaaaacacagcttaaaaactaagaatgaagggttaaaaactgagaat 459	
QY	310 gaagggttaaaaactgagaatgaagggttaaaaactgagaatgaagggttaaaaactgag 369	
DB		
DB	460 gaagggttaaaaactgagaatgaagggttaaaaactgagaatgaagggttaaaaactgag 519	
QY	370 gtgcacagagtggtttcttagggcggttagaaaacccgagcaagacacgagaaactctt 429	
DB		
DB	520 gtgcacagagtggtttcttagggcggttagaaaacccgagcaagacacgagaaactctt 579	
QY	430 aacaagtatgacgtagagaactctatgtttacaagctaaatgacaaagtaccatggaga 489	
DB		
QY	580 aacaagtatgacgtagagaactctatgtttacaagctaaatgacaaagtaccatggaga 639	
QY	490 gtgcgttatactagcagatacgcagagaataaagctaaataattattgacgactcttgac 549	
DB		
DB	640 gtgcgttatactagcagatacgcagagaataaagctaaataattattgacgactcttgac 699	
QY	550 gcaaaagaacatgaattacacacacagaagaatgagaagtattct 591	
DB		
DB	700 gcaaaagaacatgaattacacacacagaagaatgagaagtattct 741	
RESULT	4	
Q45214		
ID	Q45214 standard; DNA; 861 BP.	
XX	Q45214;	
AC	Q45214;	
XX	02-NOV-1994 (first entry)	
DT	Recombinant M24-M5-M6 gene.	
DE		
XX	Primer: PCR: amplify; polymerase chain reaction; construct; hybrid;	
KW	M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;	
KW	tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;	
KW	restriction enzyme site; multivalent M protein; immunisation; group A;	
KW	streptococci; rheumatic fever; rheumatic heart disease; humoral;	
KW	antibody; heart tissue; antigen; serotype; mucosal; ss.	
XX	Synthetic.	
OS		
XX	Key Location/Qualifiers	
EH	misc_difference 781..783	
FT	/*tag= a	
FT	/codon= seq:gga, aa:Arg	
XX	WO9406421-A.	
PN	31-MAR-1994.	
XX	15-SEP-1993; 93WO-US08703.	
XX	16-SEP-1992; 92US-0945954.	
XX	(UYTE-) UNIV TENNESSEE RES CORP.	
XX	Dale JB, Lederer JW;	
PI	WPI: 1994-118122/14.	
DR	P-PSDB; R50992.	
DR	New immunogenic hybrid proteins derives from streptococcal M proteins	
PT	- induces opsonic antibodies, for protective immunisation against	
PT	against multiple group A streptococci serotypes	
XX	Disclosure; Fig 1; 67pp; English.	
XX	The sequences given in Q45214-23 encode hybrid M proteins which	
CC	contain the M24-M5-M6 and/or M19 subunits. These tetraivalent	
CC	proteins were constructed using fragments of the 5' regions of emm	
CC	genes that were amplified by PCR, ligated in tandem and expressed in	
CC	PKK223.3. The amplified regions pref. encode protective and not	
CC	tissue-cross-reactive epitopes, which can then be linked into one	
CC	protein molecule. The recombinant hybrid protein may contain 113	
CC	N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and	
CC	35 from M19. Each section is linked by 2 amino acids specified by	
CC	the respective restriction enzyme sites that were synthesised into	
CC	the primers used to specify the PCR product. Multivalent M proteins	
CC	such as this may be used for protective immunisation against group A	
CC	streptococci, which esp. cause rheumatic fever and rheumatic heart	
CC	disease. Humoral antibodies raised against these proteins do not	
CC	react with heart tissue antigens but are effective against many	
CC	different serotypes. The multivalent proteins may also include	
CC	sequences which induce mucosal antibodies and do not require coupling	
CC	to an immunogenic carrier.	
XX	Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;	
SQ		

CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.
 XX Sequence 861 BP; 331 A; 147 C; 199 G; 184 T; 0 other;

Query Match 20.7%; Score 240; DB 15; Length 861;
 Best Local Similarity 100.0%; Pred. No. 1.3e-109;
 Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 910 gtcgcgactaggtctcagacagatactctgaaaaagctacaagaagcgtgctgacaagttt 969
 Db 4 gtcgcgactaggtctcagacagatactctgaaaaagctacaagaagcgtgctgacaagttt 63
 QY 970 gagatagaaaacaatacgtttaaaacttaagaatagtgacttaagtttttaataataaagcg 1029
 Db 64 gagatagaaaacaatacgtttaaaacttaagaatagtgacttaagtttttaataataaagcg 123
 QY 1030 ttaaaagatcataatgatgagtttaactgaagaagttgagtaagtctaagaagaaactacgt 1089
 Db 124 ttaaaagatcataatgatgagtttaactgaagaagttgagtaagtctaagaagaaactacgt 183
 QY 1090 aaaaatgataaatacactatctgaaaaagctagtaaaattcaagaattagagcagcgttaag 1149
 Db 184 aaaaatgataaatacactatctgaaaaagctagtaaaattcaagaattagagcagcgttaag 243

RESULT 5
 A57893
 ID A57893 standard; DNA; 861 BP.
 XX
 AC A57893;
 XX
 DT 10-OCT-2000 (first entry)
 XX
 DE S. Pyogenes hybrid M protein (M24-M5-M6) DNA, SEQ ID NO:1.
 XX
 KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia; ds.
 XX
 OS Streptococcus pyogenes.
 OS Synthetic.
 XX
 XX US6063386-A.
 XX
 XX 16-MAY-2000.
 XX
 XX 15-SEP-1997; 97US-0937271.
 XX
 XX 16-SEP-1992; 92US-0945954.
 XX
 XX (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PA Lederer JW, Dale JB;
 PI
 XX WPI; 2000-364475/31.
 DR
 DR P-PSDB; B03113.
 XX
 XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever
 XX
 XX Disclosure; Fig 1A-B; 62pp; English.
 XX
 CC The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.

CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue, which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences A57893-A57902
 CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
 CC proteins generated in the disclosure of the invention.
 XX
 SQ Sequence 861 BP; 331 A; 147 C; 199 G; 184 T; 0 other;

Query Match 20.7%; Score 240; DB 21; Length 861;
 Best Local Similarity 100.0%; Pred. No. 1.3e-109;
 Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 910 gtcgcgactaggtctcagacagatactctgaaaaagctacaagaagcgtgctgacaagttt 969
 Db 4 gtcgcgactaggtctcagacagatactctgaaaaagctacaagaagcgtgctgacaagttt 63
 QY 970 gagatagaaaacaatacgtttaaaacttaagaatagtgacttaagtttttaataataaagcg 1029
 Db 64 gagatagaaaacaatacgtttaaaacttaagaatagtgacttaagtttttaataataaagcg 123
 QY 1030 ttaaaagatcataatgatgagtttaactgaagaagttgagtaagtctaagaagaaactacgt 1089
 Db 124 ttaaaagatcataatgatgagtttaactgaagaagttgagtaagtctaagaagaaactacgt 183
 QY 1090 aaaaatgataaatacactatctgaaaaagctagtaaaattcaagaattagagcagcgttaag 1149
 Db 184 aaaaatgataaatacactatctgaaaaagctagtaaaattcaagaattagagcagcgttaag 243

RESULT 6
 Q45220
 ID Q45220 standard; DNA; 522 BP.
 XX
 AC Q45220;
 XX
 DT 02-NOV-1994 (first entry)
 XX
 DE Recombinant M24-M5 divalent hybrid gene.
 XX
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
 KW antibody; heart tissue; antigen; serotype; mucosal; ss.
 XX
 OS Synthetic.
 XX
 XX WO9406421-A.
 PN
 XX 31-MAR-1994.
 PD
 XX 15-SEP-1993; 93WO-US08703.
 PF
 XX 16-SEP-1992; 92US-0945954.
 PR
 XX (UYTE-) UNIV TENNESSEE RES CORP.
 PA

PI Dale JB, Lederer JW;
XX WPI; 1994-118122/14.
DR P-PSDB; R50998.
XX
XX New immunogenic hybrid proteins derives from streptococcal M proteins
PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes
XX
XX Disclosure; Fig 10; 67pp; English.
XX
XX The sequences given in Q45214-23 encode hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC PK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
CC to an immunogenic carrier.
XX
XX Sequence 522 BP; 226 A; 68 C; 116 G; 112 T; 0 other;
SQ

Query Match 18.7%; Score 217; DB 15; Length 522;
Best Local Similarity 100.0%; Pred. No. 3.4e-98;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaaaagtacagaacgctgctgacaagttt 969
Db 4 gtcgcgactaggtctcagacagatactctggaaaagtacagaacgctgctgacaagttt 63

QY 970 gagatagaaaacaatacgtttaaacttaagaatagtgacttaagttaaataaagcg 1029
Db 64 gagatagaaaacaatacgtttaaacttaagaatagtgacttaagttaaataaagcg 123

QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagagaaactacgt 1089
Db 124 ttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagagaaactacgt 183

QY 1090 aaaaatgataaatacactatctgaaaaagctagtataaa 1126
Db 184 aaaaatgataaatacactatctgaaaaagctagtataaa 220

RESULT 7
A57899
ID A57899 standard; DNA; 522 BP.
XX
XX A57899;
AC
XX 10-OCN-2000 (first entry)
DT
XX
DE S. pyogenes hybrid M protein (M24-M5) DNA, SEQ ID NO:14.
XX
XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia; ds.
XX
XX Streptococcus pyogenes.
OS Synthetic.
XX

PN US0603386-A.
XX
XX 16-MAY-2000.
XX
XX 15-SEP-1997; 97US-0937271.
XX
XX 16-SEP-1992; 92US-0945954.
XX
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX
XX Lederer JW, Dale JB;
PI
XX
XX WPI; 2000-364475/31.
XX P-PSDB; B03120.
XX
XX New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever;
XX
XX Disclosure; Fig 10A-B; 62pp; English.
XX
XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for administering a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences A57893-A57902
CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
CC proteins generated in the disclosure of the invention.
XX
XX Sequence 522 BP; 226 A; 68 C; 115 G; 113 T; 0 other;
SQ

Query Match 18.7%; Score 217; DB 21; Length 522;
Best Local Similarity 100.0%; Pred. No. 3.4e-98;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacagatactctggaaaagtacagaacgctgctgacaagttt 969
Db 4 gtcgcgactaggtctcagacagatactctggaaaagtacagaacgctgctgacaagttt 63

QY 970 gagatagaaaacaatacgtttaaacttaagaatagtgacttaagttaaataaagcg 1029
Db 64 gagatagaaaacaatacgtttaaacttaagaatagtgacttaagttaaataaagcg 123

QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagagaaactacgt 1089
Db 124 ttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagagaaactacgt 183

QY 1090 aaaaatgataaatacactatctgaaaaagctagtataaa 1126
Db 184 aaaaatgataaatacactatctgaaaaagctagtataaa 220

RESULT 8
Q45219

DB 526 ttaaaagatcataatgatgagtttaactgaagagttgagtaagtaagaaagaaactacgt 585
QY 1090 aaaaatgataaactactatctgaaaaagctagtaaaa 1126
DB 586 aaaaatgataaactactatctgaaaaagctagtaaaa 622

RESULT 9
A57898
ID A57898 standard; DNA; 741 BP.
XX
AC A57898;
XX
DT 10-OCT-2000 (first entry)
XX
DE S. pyogenes hybrid M protein (M19-M6-M5-M24) DNA, SEQ ID NO:12.
XX
KW Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
XX pneumonia; ds.
OS Streptococcus pyogenes.
OS Synthetic.
XX
PN US6063386-A.
XX
PD 16-MAY-2000.
XX
PF 15-SEP-1997; 97US-0937271.
XX
PR 16-SEP-1992; 92US-0945954.
XX
PA (UYTE-) UNIV TENNESSEE RES CORP.
XX
PI Lederer JW, Dale JB;
XX
DR WPI: 2000-364475/31.
DR P-PSDB; B03119.
XX
PT New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX
PS Disclosure; Fig 9A-B; 62pp; English.
XX
CC The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences A57893-A57902
CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
CC proteins generated in the disclosure of the invention.

Q45219 standard; DNA; 741 BP.
Q45219;
02-NOV-1994 (first entry)
Recombinant M19-M6-M5-M24 tetraivalent hybrid gene.
Primer: PCR; amplify; polymerase chain reaction; construct; hybrid;
M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
restriction enzyme site; multivalent M protein; immunisation; group A;
streptococci; rheumatic fever; rheumatic heart disease; humoral;
antibody; heart tissue; antigen; serotype; mucosal; ss.
Synthetic.
WO9406421-A.
31-MAR-1994.
15-SEP-1993; 93WO-US08703.
16-SEP-1992; 92US-0945954.
(UYTE-) UNIV TENNESSEE RES CORP.
Dale JB, Lederer JW;
WPI: 1994-118122/14.
P-PSDB; R50997.
New immunogenic hybrid proteins derives from streptococcal M proteins
- induces opsonic antibodies, for protective immunisation against
against multiple group A streptococci serotypes
Disclosure; Fig 9; 67pp; English.
The sequences given in Q45214-23 encode hybrid M proteins which
contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
proteins were constructed using fragments of the 5' regions of emm
genes that were amplified by PCR, ligated in tandem and expressed in
PKK223.3. The amplified regions pref. encode protective and not
tissue-cross-reactive epitopes, which can then be linked into one
protein molecule. The recombinant hybrid protein may contain 113
N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
35 from M19. Each section is linked by 2 amino acids specified by
the respective restriction enzyme sites that were synthesised into
the primers used to specify the PCR product. Multivalent M proteins
such as this may be used for protective immunisation against group A
streptococci, which esp. cause rheumatic fever and rheumatic heart
disease. Humoral antibodies raised against these proteins do not
react with heart tissue antigens but are effective against many
different serotypes. The multivalent proteins may also include
sequences which induce mucosal antibodies and do not require coupling
to an immunogenic carrier.
Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;

Query Match 18.7%; Score 217; DB 15; Length 741;
Best Local Similarity 100.0%; Pred. No. 3.4e-98;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 910 gtcgcgactaggtctcagacgatactctggaaaaagtacaaagctgctgacaagttt 969
DB 406 gtcgcgactaggtctcagacgatactctggaaaaagtacaaagctgctgacaagttt 465
QY 970 gagatagaaacaatacgtttaaactaaagtaagtactagtttttaataaagcg 1029
DB 466 gagatagaaacaatacgtttaaactaaagtaagtactagtttttaataaagcg 525
QY 1030 ttaaaagatcataatgatgagtttaactgaagagttgagtaagtaagaaagaaactacgt 1089

XX The sequences given in Q45214-23 encode hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These trivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC pKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC potential amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC

PS Disclosure; Fig 6A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A

CC streptococcal M proteins comprising N-terminal peptide fragments of M

CC proteins that elicit opsonic antibodies against multiple serotypes

CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies

CC generated using the hybrid proteins are against one or more M protein

CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.

CC The invention also encompasses a recombinant DNA molecule comprising a

CC nucleotide sequence that encodes a multivalent hybrid M protein; and a

CC method for immunising a mammal against streptococcal infections,

CC comprising administering an immunogenic multivalent hybrid M protein to

CC the mammal. The multivalent hybrid M proteins are useful for eliciting

CC opsonic or protective antibodies to the M proteins of different serotypes

CC of group A streptococci, and may therefore be used as vaccines to protect

CC against and control infection by type A streptococci. Type A streptococci

CC are not only responsible for streptococcal pharyngitis (strep throat),

CC forms of pneumonia and a condition resembling toxic shock, but are also

CC involved in the development of acute rheumatic fever (ARF) and rheumatic

CC heart disease. In a patient with ARF, antibodies formed during a group A

CC streptococcal infection are also cross-reactive with heart tissue, which

CC indicates that the streptococci and host tissue contain similar antigenic

CC motifs. The new multivalent vaccines are capable of raising sero-specific

CC antibodies against various serotypes of group A streptococci which are

CC not cross-reactive with human heart tissue. Sequences A57893-A57902

CC represent DNAs encoding multivalent hybrid Streptococcus pyogenes M

CC proteins generated in the disclosure of the invention.

XX

XX Sequence 822 BP; 332 A; 136 C; 182 G; 172 T; 0 other;

SQ

Query Match 18.7%; Score 217; DB 21; Length 822;

Best Local Similarity 100.0%; Pred. No. 3.4e-98;

Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 910 gtgcgactaggtctcaagacagatactctggaaaaagatcacagaagctgtcgaagttt 969

Db 4 gtccgactaggtctccagacagatactctggaaaaagatcacagaagctgtcgaagttt 63

Qy 970 gagatgaaaaacaatactctaaaaacttaagaatagtgacttaagttttaataaagcg 1029

Db 64 gagatgaaaaacaatactctaaaaacttaagaatagtgacttaagttttaataaagcg 123

Qy 1030 ttaaaagatcataatgatgagtgtaactgaagagtgtagtaagtctaaagaaactacgt 1089

Db 124 ttaaaagatcataatgatgagtgtaactgaagagtgtagtaagtctaaagaaactacgt 183

Qy 1090 aaaaatgataaatcactatctgaaaaagcgtagtaaaa 1126

Db 184 aaaaatgataaatcactatctgaaaaagcgtagtaaaa 220

RESULT 12

Q45221

ID Q45221 standard; DNA; 1029 BP.

XX

XX Q45221;

AC

XX

XX 02-NOV-1994 (first entry)

DT

XX

DE Tetravalent-C repeat gene.

XX

KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;

KW M protein; M24; M5; M6; M19; subunit; tetravalent protein; emm gene;

KW tandem; pK223.3; protective epitope; tissue-cross-reactive epitope;

KW restriction enzyme site; multivalent M protein; immunisation; group A;

KW streptococci; rheumatic fever; rheumatic heart disease; humoral;

KW antibody; heart tissue; antigen; serotype; mucosal; ss.

XX

OS Synthetic.

XX

XX S. pyogenes hybrid M protein (M19-M6-M5-M24) DNA, SEQ ID NO:15.

PN W09406421-A.

XX

PD 31-MAR-1994.

XX

PF 15-SEP-1993; 93WO-US08703.

XX

PR 16-SEP-1992; 92US-0945954.

XX

PA (UYTE-) UNIV TENNESSEE RES CORP.

XX

PI Dale JB, Lederer JW;

XX

XX WPI: 1994-118122/14.

DR P-PSDB: R50999.

XX

XX New immunogenic hybrid proteins derives from streptococcal M proteins

PT - induces opsonic antibodies, for protective immunisation against

PT against multiple group A streptococci serotypes

XX

PS Disclosure; Fig 11; 67pp; English.

XX

CC The sequences given in Q45214-23 encode hybrid M proteins which

CC contain the M24-M5-M6 and/or M19 subunits. These tetravalent

CC proteins were constructed using fragments of the 5' regions of emm

CC genes that were amplified by PCR, ligated in tandem and expressed in

CC pK223.3. The amplified regions pref. encode protective and not

CC tissue-cross-reactive epitopes, which can then be linked into one

CC protein molecule. The recombinant hybrid protein may contain 113

CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and

CC 35 from M19. Each section is linked by 2 amino acids specified by

CC the respective restriction enzyme sites that were synthesised into

CC the primers used to specify the PCR product. Multivalent M proteins

CC such as this may be used for protective immunisation against group A

CC streptococci, which esp. cause rheumatic fever and rheumatic heart

CC disease. Humoral antibodies raised against these proteins do not

CC react with heart tissue antigens but are effective against many

CC different serotypes. The multivalent proteins may also include

CC sequences which induce mucosal antibodies and do not require coupling

CC to an immunogenic carrier.

SQ

Sequence 1029 BP; 424 A; 168 C; 226 G; 211 T; 0 other;

Query Match 18.7%; Score 217; DB 15; Length 1029;

Best Local Similarity 100.0%; Pred. No. 3.4e-98;

Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 910 gtccgactaggtctcaagacagatactctggaaaaagatcacagaagctgtcgaagttt 969

Db 406 gtccgactaggtctccagacagatactctggaaaaagatcacagaagctgtcgaagttt 465

Qy 970 gagatgaaaaacaatactctaaaaacttaagaatagtgacttaagttttaataaagcg 1029

Db 466 gagatgaaaaacaatactctaaaaacttaagaatagtgacttaagttttaataaagcg 525

Qy 1030 ttaaaagatcataatgatgagtgtaactgaagagtgtagtaagtctaaagaaactacgt 1089

Db 526 ttaaaagatcataatgatgagtgtaactgaagagtgtagtaagtctaaagaaactacgt 595

Qy 1090 aaaaatgataaatcactatctgaaaaagcgtagtaaaa 1126

Db 586 aaaaatgataaatcactatctgaaaaagcgtagtaaaa 622

RESULT 13

A57900

ID A57900 standard; DNA; 1029 BP.

XX

XX A57900;

AC

XX

XX 10-OCT-2000 (first entry)

DT

XX

DE S. pyogenes hybrid M protein (M19-M6-M5-M24) DNA, SEQ ID NO:15.

XX

KW Multivalent hybrid M protein; group A streptococcus; serotype;

Tue May 15 07:26:45 2001

immunogenic; sero-specific antibody; streptococcal infection;
cross reactivity; vaccine; acute rheumatic fever; ARF;
rheumatic heart disease; streptococcal pharyngitis; strep throat;
pneumonia; ds.

Streptococcus pyogenes.

Synthetic.

US6063386-A.

16-MAY-2000.

15-SEP-1997; 97US-0937271.

16-SEP-1992; 92US-0945954.

(UYTE-) UNIV TENNESSEE RES CORP.

Lederer JW, Dale JB;

WPI; 2000-364475/31.

P-PSDB; B03121.

New immunogenic recombinant hybrid M protein comprising amino-terminal
peptide fragments of streptococcal M protein useful as vaccine against
rheumatic fever and infections leading to rheumatic fever

Disclosure; Fig 11A-C; 62pp; English.

The invention relates to multivalent immunogenic hybrid group A
streptococcal M proteins comprising N-terminal peptide fragments of M
proteins that elicit opsonic antibodies against multiple serotypes
of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
generated using the hybrid proteins are against one or more M protein
serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
The invention also encompasses a recombinant DNA molecule comprising a
nucleotide sequence that encodes a multivalent hybrid M protein; and a
method for immunising a mammal against streptococcal infections,
comprising administering an immunogenic multivalent hybrid M protein to
the mammal. The multivalent hybrid M proteins are useful for eliciting
opsonic or protective antibodies to the M proteins of different serotypes
of group A streptococci, and may therefore be used as vaccines to protect
against and control infection by type A streptococci. Type A streptococci
are not only responsible for streptococcal pharyngitis (strep throat),
but are also involved in the development of acute rheumatic fever (ARF) and rheumatic
heart disease. In a patient with ARF, antibodies formed during a group A
streptococcal infection are also cross-reactive with heart tissue, which
indicates that the streptococci and host tissue contain similar antigenic
motifs. The new multivalent vaccines are capable of raising sero-specific
antibodies against various serotypes of group A streptococci which are
not cross-reactive with human heart tissue. Sequences A57893-A57902
represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
proteins generated in the disclosure of the invention.

Sequence 1029 BP; 424 A; 168 C; 226 G; 211 T; 0 other;

Query Match 18.7%; Score 217; DB 21; Length 1029;
Best Local Similarity 100.0%; Pred. No. 3.4e-98;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 910 gtcgcactaggtctcagacagatactctggaataagctgacttaagttttaataataaagcg 1029
Db 406 gtcgcactaggtctcagacagatactctggaataagctgacttaagttttaataataaagcg 525
Qy 970 gagatagaataacacgtttaaaacttaagaataagctgacttaagttttaataataaagcg 1089
Db 466 gagatagaataacacgtttaaaacttaagaataagctgacttaagttttaataataaagcg 525
Qy 1030 ttaaaagatcataatgatgatgttaactgaaggttgagtaagtctaaagaaactacgt 1089
Db 526 ttaaaagatcataatgatgatgttaactgaaggttgagtaagtctaaagaaactacgt 585

Qy 1090 aaaaatgataaatacactatctgtgaaaaagctagtaaaa 1126
Db 586 aaaaatgataaatacactatctgtgaaaaagctagtaaaa 622

RESULT 14

T49317
ID T49317 standard; DNA; 1617 BP.

XX
AC T49317;

XX
DT 31-MAR-1997 (first entry)

XX
DE Type-6 M-protein emm6 gene.

XX
KW Type-6 M-protein; emm6 gene; Streptococcus pyogenes; signal peptide;
anchor peptide; C-terminal sorting signal; surface display;
fusion protein; protease-deficient; Streptococcus gordonii;
KW SPFX vector; fusion protein cleavage; spacer; protein secretion;
KW TEV-N1a protease; diagnostic; recombinant vaccine; therapy; ss.

OS Streptococcus pyogenes.

XX
FH Key Location/Qualifiers

FT RBS 87..93

FT sig_peptide /*tag= a

FT /*tag= b

FT CDS 100..1551

FT /*tag= c

FT mat_peptide /product= Type-6 M-protein

FT terminator /*tag= d

FT /*tag= e

FT /rpt_type= INVERTED

XX
PN W09640943-A1.

XX
PD 19-DEC-1996.

XX
PE 06-JUN-1996; 96WO-US09965.

XX
PX 07-JUN-1995; 95US-0472244.

XX
PA (UYRQ) UNIV ROCKEFELLER.

XX
PI Darzins A, Hruby D, Whitehead S;

XX
PX WPI; 1997-052336/05.

XX
DR P-PSDB; W08927.

XX
PT Expressing heterologous proteins in gram-positive bacteria -
PT Produces proteins anchored to host's cell surface, or secreted from
PT cell

XX
PS Disclosure; Fig 3; 41pp; English.

XX
CC This sequence encodes Streptococcus pyogenes type-6 M-protein,
CC which is a cell wall surface protein. The N-terminal signal
CC peptide, part of the mature N-terminal sequence, the C-terminal
CC sorting signal (anchor) sequence and optionally a spacer region
CC from the protein may be expressed as a fusion with a target protein
CC in a protease-deficient gram-positive host, e.g. the human oral
CC commensal Streptococcus gordonii, in a new cloning method involving
CC the SPFX vector series (e.g. plasmid pSPFX1a). The resulting fusion
CC protein is expressed on the cell surface, and may be cleaved with a
CC protease (at a site adjacent to the anchor) for simple purification.
CC A vector without the anchor sequence may be used for protein
CC secretion. The method may be used in production of recombinant
CC diagnostic, vaccine and therapeutic proteins.

XX

us-09-151-409-15.oligo.rng

Tue May 15 07:26:45 2001

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
; US-08-937-271-3

Query Match          34.7%; Score 402; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 5,le-195;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 190 ggtccgcgcctgactagggtacaataatgcccgcaaaagagcaaaagaagcttcttac 249
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 340 GGATCGCCGCTGACTAGGGGTACATAAATGACCCTCAAGAGCAAAAGAACCTCTTGAC 399
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 250 aagttagctagtagaaacaccatgactctaaaaaactaaagaatgaagggttaaaaactgagaat 309
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 400 AAGTATGAGCTAGAAAAACCATGACTTTAAAACCTTAAGAATCAAGGGTTAAAAACTGAGAA 459
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 310 gaagggttaaaaactgagaatgaagggttaaagaactgagaatgaagggttaaaaactgag 369
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 460 GAAGGGTTAAAAACTGAGAATGAAGGGTTAAAAACTGAGAAATGAAGGGTTAAAAACTGAG 519
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 370 gtccagacagtgtttccttagggggaoggtagaiaaacccggagcaaacacagagaacttctt 429
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 520 CTCACACAGAGTGTTCCTTAGGGGCAGCGTAGAAAACCCGCCAGAACACGACGAGAACTTCTT 579
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 430 acaaagtatgcgtagagaactctatgttacaagctaataatgacaagttaccatggaga 489
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 580 AACAAAGTATGACGTAGAGAACTCTATGTTACNAGCTAATAATGACAAGTTACCATTGGAGA 639
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 490 gtgcgttatactaggcatcacccgaagagataagctaaaaaattattgcgatcttgac 549
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 640 GTGCGTTTACTAGGCATACGCCAGAGATGAAGCTAAAAAAAATTAITGACGATCTTGAC 699
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 550 gcaaaagacatgaattacaacaacagaaatgagaagtattct 591
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 700 GC AAAAGAACATGAATTAACAACACAGAAATCAGAAGTTATCT 741
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT      2
US-08-937-271-1
; Sequence 1, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; DATE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
```

SEQUENCE CHARACTERISTICS:
LENGTH: 861 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..861
US-08-937-271-1

Query Match 20.7% Score 240; DB 3; Length 861;
Best Local Similarity 100.0%; Pred. No. 9.5e-113; Mismatches 0; Indels 0; Gaps 0;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 910 gtcgcgactaggtctcagacagatactctggaaaaagttacaaagcgtgctgacaagttt 969
Db 4 GTCGCGACTAGGTCTCAGACAGATACTCTGGAAAAAGTACAAGACGTGCTGACAAGTTT 63
QY 970 gagatagaacaatacgtttaaaacttaagaatagtgacttaagttttaataataaaagcg 1029
Db 64 GAGATAGAAAAACAATACGTTAAACTTAAGATAGTAGTGAAGTTTAAATAATAAAGCG 123
QY 1030 ttaaaagatcataatgatagtttaaaactgaagagttgagtaactgctaaagagaaactacgt 1089
Db 124 TTAAGATCATATGATGAGTTAACTGAAGAGTTTCAGTAAATGCTAAAGAGAACTAGCT 183
QY 1090 aaaaatgataaatcactctctgaaaaagctagtaaaatcaagaattagagcgacgtaag 1149
Db 184 AAAAATGATAAATCACTACTCTGAAAANGCTAGTAAATTAAGAATTAGAGGCACGTAAG 243

RESULT 3
US-08-937-271-14
; Sequence 14, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..519
US-08-937-271-14

Query Match 18.7% Score 217; DB 3; Length 522;
Best Local Similarity 100.0%; Pred. No. 4.6e-101; Mismatches 0; Indels 0; Gaps 0;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 910 gtcgcgactaggtctcagacagatactctggaaaaagttacaaagcgtgctgacaagttt 969
Db 4 GTCGCGACTAGGTCTCAGACAGATACTCTGGAAAAAGTACAAGACGTGCTGACAAGTTT 63
QY 970 gagatagaacaatacgtttaaaacttaagaatagtgacttaagttttaataataaaagcg 1029
Db 64 GAGATAGAAAAACAATACGTTAAACTTAAGATAGTAGTGAAGTTTAAATAATAAAGCG 123
QY 1030 ttaaaagatcataatgatagtttaaaactgaagagttgagtaactgctaaagagaaactacgt 1089
Db 124 TTAAGATCATATGATGAGTTAACTGAAGAGTTTCAGTAAATGCTAAAGAGAACTAGCT 183
QY 1090 aaaaatgataaatcactctctgaaaaagctagtaaaatcaagaattagagcgacgtaag 1126
Db 184 AAAAATGATAAATCACTACTCTGAAAANGCTAGTAAATTAAGAATTAGAGGCACGTAAG 220

RESULT 4
US-08-937-271-12
; Sequence 12, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 741 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

us-09-151-409-15.oligo.rni

Tue May 15 07:26:50 2001

```

; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
; US-08-937-271-12

Query Match      18.7%; Score 217; DB 3; Length 741;
Best Local Similarity 100.0%; Pred. No. 4.6e-101; Indels 0; Gaps 0;
Matches 217; Conservative 0; Mismatches 0;

QY 910 gtcgcgactaggtctcagacagatactctggaataacttaagaatagtgacttaagttttaataaagcg 1029
Db 406 GTCCGACTAGTCTCAGACAGATACTCTGGAAAAGTACAAAGACGTGCTGACAAGTTT 465
QY 970 gagatgagaataacttcgtaaaacttaagaatagtgacttaagttttaataaagcg 1029
Db 466 GAGATGAGAAAACAAATACGTTAAACCTTAAGATATAGTGAATTTAAATATAAAGCG 525
QY 1030 ttaaaagatcataatgatgagtttaactgaagagtgagtaagtctaaagagaaactacgt 1089
Db 526 TTAAGATGATCAATGATGAGTTAACTGAAGAGTTGAGTAATGCTTAAAGAGAAACTACGT 585
QY 1090 aaaaatgataatcaactatctgaaaagctagtaaaa 1126
Db 586 AAAAATGATAAATCACTATCTGAAAAGCTAGTAAAA 622

RESULT 5
US-08-937-271-5
; Sequence 5, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; HYPOTHETICAL: NO
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

; TOPOLOGY: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..822
; US-08-937-271-5

Query Match      18.7%; Score 217; DB 3; Length 822;
Best Local Similarity 100.0%; Pred. No. 4.6e-101; Indels 0; Gaps 0;
Matches 217; Conservative 0; Mismatches 0;

QY 910 gtcgcgactaggtctcagacagatactctggaataacttaagaatagtgacttaagttttaataaagcg 1029
Db 4 GTCCGACTAGTCTCAGACAGATACTCTGGAAAAGTACAAAGACGTGCTGACAAGTTT 63
QY 970 gagatgagaataacttcgtaaaacttaagaatagtgacttaagttttaataaagcg 1029
Db 64 GAGATGAGAAAACAAATACGTTAAACCTTAAGATATAGTGAATTTAAATATAAAGCG 123
QY 1030 ttaaaagatcataatgatgagtttaactgaagagtgagtaagtctaaagagaaactacgt 1089
Db 124 TTAAGATGATCAATGATGAGTTAACTGAAGAGTTGAGTAATGCTTAAAGAGAAACTACGT 183
QY 1090 aaaaatgataatcaactatctgaaaagctagtaaaa 1126
Db 184 AAAAATGATAAATCACTATCTGAAAAGCTAGTAAAA 220

RESULT 6
US-08-937-271-16
; Sequence 16, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1029
US-08-937-271-16

Query Match 18.7%; Score 217; DB 3; Length 1029;
Best Local Similarity 100.0%; Pred. No. 4.5e-101;
Matches 217; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 910 gtcgcgactaggtctcagacagatctctggaaaaaatacaaaagtcgtgacaagttt 969
Db 406 GTCGCGACTAGTCTCAGACAGATCTCTGAAAAAGTACAAGAAGCTGTGACAAGTTT
QY 970 gagatagaaaaatacgtttaaaacttaagaatagtgacttaagttttaataataaagcg 1029
Db 466 GAGATAGAAAAAATACTGTTAAAACTTAAGAATAGTACTTAAGTTTATAATAAAGCG 525
QY 1030 ttaaaagatcataatgatgtgagtttaactgaagtgagtaagtctaaagagaaactacgt 1089
Db 526 TTAAGAATGATCAATGATGATGATTAACCTGAAGAGTTGAGTAATGCTAAAGAGAACTACCT 585
QY 1090 aaaaatgataatcactatctgaaaaagctagtataaa 1126
Db 586 AAAAATGATAATCACTATCTGAAAAAGCTAGTAAAA 622

RESULT 7
US-08-795-475-5
; Sequence 5, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; APPLICANT: Sjvdring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795.475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1329
US-08-795-475-5

Query Match 13.0%; Score 151; DB 2; Length 1332;
Best Local Similarity 100.0%; Pred. No. 1.5e-67;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 aacggatggttaactcctaggaagttatagaagatcttcagcaaaacatcccgaata 657
Db 1 AACGGTGATGTAATCTCTAGGAAAGTTTATAGAAGATCTTCAGCAAAACAATCCCGCAATA 60
QY 658 caaatatatacgtttacgtcagcaaaacaaagactttaaagcgagattagagaatgcaatg 717
Db 61 CAATATATACGTTTACGTCTCAGAAAACAAGGACTTAAAGCGAGATTAGAGATGCAATG 120
QY 718 gaatttcaggaagagagattttaagagagctg 748
Db 121 GAAGTTGAGCAAGAGAGATTTTAAAGAGAGCTG 151
RESULT 8
US-08-937-271-7
; Sequence 7, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937.271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..561
US-08-937-271-7

Query Match 4.6%; Score 53; DB 3; Length 561;
Best Local Similarity 100.0%; Pred. No. 9e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 11
US-08-937-271-9
; Sequence 9, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```


APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 918 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..915
US-08-937-271-9

Query Match 4.5%; Score 52; DB 3; Length 918;
Best Local Similarity 100.0%; Pred. No. 2.8e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 190 ggatccgcctgactaggggtacataaatgacccgcaagagcaagaag 241
|||||
Db 49 GCATCCGCGTGACTAGGGGTACATAATAATGACCCGCAAGACCAAGAAAG 100

RESULT 12
US-08-937-271-26
Sequence 26, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
US-08-937-271-26

Query Match 3.3%; Score 38; DB 3; Length 39;
Best Local Similarity 100.0%; Pred. No. 4.1e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 191 gatccgcctgactaggggtacataaatgacccgcaaa 228
|||||
Db 1 GATCCGCGCTGACTAGGGGTACATAATAATGACCCGCAA 38

RESULT 13
US-08-937-271-28
Sequence 28, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-937-271-28

Query Match 2.9%; Score 34; DB 3; Length 47;
Best Local Similarity 100.0%; Pred. No. 4.3e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 371 tcgacagagtgttctcctaggggacgtagaaaa 404
|||||
Db 1 TCGACAGAGTGTTCCTAGGGGACGCTAGAAAA 34

RESULT 14
US-08-937-271-38
Sequence 38, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN

Tue May 15 07:26:50 2001

;; TITLE OF INVENTION: VACCINE
;; NUMBER OF SEQUENCES: 40
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SEED AND BERRY
;; STREET: 6300 Columbia Center, 701 Fifth Avenue
;; CITY: Seattle
;; STATE: Washington
;; COUNTRY: USA
;; ZIP: 98104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/937,271
;; FILING DATE: 15-SEP-1997
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rosenman, Stephen J.
;; REGISTRATION NUMBER: 43,058
;; REFERENCE/DOCKET NUMBER: 481112.405C1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 38:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 54 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-937-271-38

Query Match 2.8%; Score 32; DB 3; Length 54;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 491 tgcgttatacaggcaccgagaagataag 522
|||||
DB 5 TGCCTTACTAGCATACGCCAGAGATAAG 36

RESULT 15
US-08-937-271-34
;; Sequence 34, Application US/08937271
;; Patent No. 6063386
;; GENERAL INFORMATION:
;; APPLICANT: Dale, James B.
;; APPLICANT: Lederer, James W.
;; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
;; TITLE OF INVENTION: VACCINE
;; NUMBER OF SEQUENCES: 40
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: SEED AND BERRY
;; STREET: 6300 Columbia Center, 701 Fifth Avenue
;; CITY: Seattle
;; STATE: Washington
;; COUNTRY: USA
;; ZIP: 98104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/937,271
;; FILING DATE: 15-SEP-1997
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rosenman, Stephen J.
;; REGISTRATION NUMBER: 43,058
;; REFERENCE/DOCKET NUMBER: 481112.405C1

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 34:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 30 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-937-271-34

Query Match 2.2%; Score 25; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 ggtcgacagagtggtttcctaggggg 393
|||||
DB 6 GGTCCACACAGTGTTCCTAGGGGG 30

Search completed: May 13, 2001, 06:16:14
Job time: 4334 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:55:40 ; Search time 1035.91 Seconds
(without alignments)
9765.715 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcacgatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 9623517 seqs, 4368049070 residues

Word size : 10

Total number of hits satisfying chosen parameters: 6063379

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

EST:*

1: gb_estl:*

2: gb_est2:*

3: gb_est3:*

4: gb_est4:*

5: gb_est5:*

6: gb_est6:*

7: gb_est7:*

8: gb_est8:*

9: gb_est9:*

10: gb_est10:*

11: gb_est11:*

12: gb_est12:*

13: gb_est13:*

14: gb_est14:*

15: gb_est15:*

16: gb_est16:*

17: gb_est17:*

18: gb_est18:*

19: gb_est19:*

20: gb_est20:*

21: gb_est21:*

22: gb_est22:*

23: gb_est23:*

24: gb_est24:*

25: gb_est33:*

26: gb_est34:*

27: gb_est35:*

28: gb_est36:*

29: gb_est37:*

30: gb_est38:*

31: gb_est39:*

32: gb_est40:*

33: em_estba:*

34: em_estfun:*

35: em_esthum1:*

36: em_esthum2:*

37: em_esthum3:*

38: em_esthum4:*

39: em_esthum5:*

40: em_esthum6:*

41: em_esthum7:*

42: em_esthum8:*

43: em_esthum9:*

44: em_esthum10:*

45: em_esthum11:*

46: em_esthum12:*

47: em_esthum13:*

48: em_esthum14:*

49: em_esthum15:*

50: em_esthum16:*

51: em_esthum17:*

52: em_esthum18:*

53: em_esthum19:*

54: em_esthum20:*

55: em_esthum21:*

56: em_esthum22:*

57: em_esthum23:*

58: em_esthum24:*

59: em_esthum25:*

60: em_esthum26:*

61: em_esthum27:*

62: em_esthum28:*

63: em_estin1:*

64: em_estin2:*

65: em_estin3:*

66: em_estin4:*

67: em_estin5:*

68: em_estom1:*

69: em_estom2:*

70: em_estov1:*

71: em_estov2:*

72: em_estpl1:*

73: em_estpl2:*

74: em_estpl3:*

75: em_estpl4:*

76: em_estpl5:*

77: em_estpl6:*

78: em_estpl7:*

79: em_estpl8:*

80: em_estpl9:*

81: em_estpl10:*

82: em_estol:*

83: em_estol2:*

84: em_estol3:*

85: em_estol4:*

86: em_estol5:*

87: em_estol6:*

88: em_estol7:*

89: em_estol8:*

90: em_estol9:*

91: em_estol10:*

92: em_estol11:*

93: em_estol12:*

94: em_estol13:*

95: em_estol14:*

96: em_estol15:*

97: em_estol16:*

98: em_estol17:*

99: em_estol18:*

100: em_estol19:*

101: em_estol20:*

102: gb_est25:*

103: gb_est26:*

104: gb_est27:*

105: gb_est28:*

106: gb_est29:*

107: gb_est30:*

108: gb_est31:*

109: gb_est32:*

110: gb_est41:*

111: gb_est42:*

112: gb_est43:*

113: gb_est44:*

114: gb_est45:*

115: gb_est46:*

116: gb_est47:*

```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

```

```

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB ID	Description
	Score	Match	Length		
C 1	20	1.7	188	104	AI971561
C 2	20	1.7	453	112	AW229282
C 3	20	1.7	458	203	AO165852
C 4	20	1.7	465	104	AI970049
C 5	20	1.7	604	1	AA013829
C 6	20	1.7	608	165	BE311177
C 7	20	1.7	636	18	AI322463
C 8	20	1.7	665	13	AA880933
C 9	20	1.7	842	226	AZ693348
C 10	20	1.7	893	225	AZ675337
C 11	19	1.6	175	8	AA492011
C 12	19	1.6	272	147	BF362820
C 13	19	1.6	291	109	AV136611
C 14	19	1.6	331	152	HI8381
C 15	19	1.6	365	23	AI638554
C 16	19	1.6	378	169	BF832653
C 17	19	1.6	415	136	BE576549
C 18	19	1.6	419	206	AQ401806
					AI971561 wq87d08.x
					AW229282 uo08g01.y
					AQ165852 HS_3070.A
					AI970049 wq80d04.x
					AA013829 mh04h05.r
					BE311177 601057718
					AI322463 mh04h05.y
					AA880933 vx34805.f
					AZ693348 ENTMIJ05FF
					AZ675337 ENTIMW55TR
					AA492011 ne72h01.s
					BF362820 CM4-NN007
					AV136611 AV136611
					HI8381 yn49q01.r1
					AI638554 ts50a01.x
					BF832653 PM0-WT091
					BE576549 Gc41607.y
					AQ401806 HS_5045_A

C 749 16 1.4 285 108 AV064405
C 750 16 1.4 285 122 BB011370
C 751 16 1.4 285 127 BB182160
C 752 16 1.4 285 129 BB262899
C 753 16 1.4 285 134 BB483586
C 754 16 1.4 285 135 BB526826
C 755 16 1.4 285 163 BB135579
C 756 16 1.4 286 29 AV341178
C 757 16 1.4 286 29 AV341178
C 758 16 1.4 286 108 AV055022
C 759 16 1.4 286 108 AV064534
C 760 16 1.4 286 128 BB237874
C 761 16 1.4 286 134 BB472722
C 762 16 1.4 286 135 BB530035
C 763 16 1.4 286 135 BB530035
C 764 16 1.4 287 108 AV064395
C 765 16 1.4 287 124 BB067542
C 766 16 1.4 287 124 BB067928
C 767 16 1.4 287 156 T00761
C 768 16 1.4 287 166 BE352965
C 769 16 1.4 287 170 BF859017
C 770 16 1.4 288 108 AV064747
C 771 16 1.4 288 122 BB013377
C 772 16 1.4 288 131 BB366527
C 773 16 1.4 288 134 BB472577
C 774 16 1.4 288 135 BB514947
C 775 16 1.4 288 137 BE566226
C 776 16 1.4 289 108 AV061939
C 777 16 1.4 289 108 AV068977
C 778 16 1.4 289 127 BB190706
C 779 16 1.4 289 129 BB281643
C 780 16 1.4 289 134 BB485273
C 781 16 1.4 289 168 BF722010
C 782 16 1.4 290 108 AV054020
C 783 16 1.4 290 123 BB040286
C 784 16 1.4 290 129 BB282448
C 785 16 1.4 291 108 AV053774
C 786 16 1.4 291 108 AV064530
C 787 16 1.4 291 111 BB134675
C 788 16 1.4 291 128 BB322762
C 789 16 1.4 292 26 AV235830
C 790 16 1.4 292 108 AV052218
C 791 16 1.4 292 108 AV052334
C 792 16 1.4 292 108 AV064683
C 793 16 1.4 292 129 BB261124
C 794 16 1.4 292 131 BB359551
C 795 16 1.4 292 133 BB431111
C 796 16 1.4 292 133 BB439831
C 797 16 1.4 292 201 AQ012580
C 798 16 1.4 292 233 TA87A10P
C 799 16 1.4 293 108 AV058312
C 800 16 1.4 293 108 AV069973
C 801 16 1.4 293 122 BB006511
C 802 16 1.4 293 124 BB068374
C 803 16 1.4 294 27 AV271396
C 804 16 1.4 294 108 AV056321
C 805 16 1.4 294 108 AV064754
C 806 16 1.4 294 127 BB212316
C 807 16 1.4 294 129 BB263048
C 808 16 1.4 294 129 BB278863
C 809 16 1.4 294 134 BB478387
C 810 16 1.4 295 28 AV339617
C 811 16 1.4 295 108 AV064288
C 812 16 1.4 295 126 BB136278
C 813 16 1.4 295 128 BB231388
C 814 16 1.4 296 108 AV056622
C 815 16 1.4 296 129 BB282447
C 816 16 1.4 296 138 BE686345
C 817 16 1.4 296 153 R04249
C 818 16 1.4 296 165 BE136668
C 819 16 1.4 296 227 B79407
C 820 16 1.4 297 108 AV070571
C 821 16 1.4 297 134 BB471010

C 822 16 1.4 298 108 AV051971
C 823 16 1.4 298 108 AV055939
C 824 16 1.4 298 108 AV069763
C 825 16 1.4 298 124 BB083147
C 826 16 1.4 298 128 BB248536
C 827 16 1.4 298 130 BB307259
C 828 16 1.4 299 108 AV064587
C 829 16 1.4 299 123 BB040867
C 830 16 1.4 299 128 BB221106
C 831 16 1.4 299 129 BB280943
C 832 16 1.4 300 13 AA884595
C 833 16 1.4 300 124 BB079779
C 834 16 1.4 300 175 C34491
C 835 16 1.4 300 201 AQ026969
C 836 16 1.4 301 108 AV069888
C 837 16 1.4 301 108 AV080837
C 838 16 1.4 301 124 BB065023
C 839 16 1.4 301 129 BB282508
C 840 16 1.4 302 132 BB385814
C 841 16 1.4 302 134 BB069221
C 842 16 1.4 302 168 BF773154
C 843 16 1.4 302 169 BF826671
C 844 16 1.4 303 9 AA614866
C 845 16 1.4 303 14 AA969986
C 846 16 1.4 303 125 BB125963
C 847 16 1.4 303 128 BB219076
C 848 16 1.4 303 135 BB512702
C 849 16 1.4 303 232 PT011K050
C 850 16 1.4 304 28 AV338736
C 851 16 1.4 304 108 AV049996
C 852 16 1.4 304 127 BB201776
C 853 16 1.4 304 128 BB236239
C 854 16 1.4 304 134 BB476497
C 855 16 1.4 305 25 AV172102
C 856 16 1.4 305 122 AW997962
C 857 16 1.4 305 202 AQ095052
C 858 16 1.4 305 127 BB215292
C 859 16 1.4 306 128 BB217220
C 860 16 1.4 306 160 BB546658
C 861 16 1.4 306 160 BB554841
C 862 16 1.4 307 5 AA324818
C 863 16 1.4 307 7 AA615151
C 864 16 1.4 307 125 BB106774
C 865 16 1.4 307 126 BB136858
C 866 16 1.4 307 129 BB285692
C 867 16 1.4 307 132 BB393873
C 868 16 1.4 308 123 BB044192
C 869 16 1.4 308 125 BB111667
C 870 16 1.4 308 129 BB258239
C 871 16 1.4 308 160 BB549577
C 872 16 1.4 309 127 BB188699
C 873 16 1.4 309 127 BB192603
C 874 16 1.4 310 123 BB044285
C 875 16 1.4 311 127 BB206566
C 876 16 1.4 311 127 BB208293
C 877 16 1.4 311 202 AQ096307
C 878 16 1.4 312 125 BB124727
C 879 16 1.4 312 130 BB369025
C 880 16 1.4 312 160 BB542429
C 881 16 1.4 313 13 AA869173
C 882 16 1.4 313 125 BB130249
C 883 16 1.4 313 128 BB243631
C 884 16 1.4 313 129 BB281362
C 885 16 1.4 313 215 AZ042211
C 886 16 1.4 314 6 AA370281
C 887 16 1.4 314 13 AA871432
C 888 16 1.4 314 124 BB070959
C 889 16 1.4 314 130 BB200398
C 890 16 1.4 314 137 BB327409
C 891 16 1.4 314 130 BB327409
C 892 16 1.4 315 111 AW145737
C 893 16 1.4 315 111 AW145737
C 894 16 1.4 315 134 BB460179

AV051971 AV051971
AV055939 AV055939
AV069763 AV069763
BB083147 BB083147
BB248536 BB248536
BB307259 BB307259
AV064587 AV064587
BB040867 BB040867
BB221106 BB221106
BB280943 BB280943
AA884595 am38e07.s
BB079779 BB079779
C34491 C34491 YU11
AQ026969 CIT-HSP-2
AV069888 AV069888
AV080837 AV080837
BB065023 BB065023
BB282508 BB282508
BB385814 BB385814
BB069221 BB069221
BB346658 BB346658
BF773154 CM4-T004
BF826671 CM2-RT000
AA614866 VO32905.r
AA969986 op61b02.s
BB125963 BB125963
BB219076 BB219076
BB512702 BB512702
AL448488 ParamecLu
AV338736 AV338736
AV049996 AV049996
BB201776 BB201776
BB236239 BB236239
BB476497 BB476497
AV172102 AV172102
AW997962 RC1-BN005
AQ095052 HS 3027.B
BB215292 BB215292
BB217220 BB217220
BB546658 BB546658
BB554841 BB554841
AA324818 EST27735
AA615151 ZX60C03.s
BB106774 BB106774
BB136858 BB136858
BB285692 BB285692
BB393873 BB393873
BB044192 BB044192
BB111667 BB111667
BB258239 BB258239
BB549577 BB549577
BB188699 BB188699
BB192603 BB192603
BB044285 BB044285
BB206566 BB206566
BB208293 BB208293
AQ096307 HS 3030.A
BB124727 BB124727
BB369025 BB369025
BB542429 BB542429
AA869173 vq48d11.r
BB130249 BB130249
BB243631 BB243631
BB281362 BB281362
AZ042211 RFC1-23-3
AA370281 EST82141
AA871432 vq35C08.r
BB070959 BB070959
BB200398 BB200398
BB327409 BB327409
AZ321290 1M0041N20
AW145737 qa34b08.Y
BB460179 BB460179


```

source
1. .188
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2479023"
/clone_lib="NCL_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: p773D-pac (Pharmacia) with a modified
polylinker; Plasmid DNA from the normalized library
NCL_CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (cloneIDs 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT      65 a 52 c 16 g 55 t
ORIGIN

Query Match      1.7%: Score 20; DB 104; Length 188;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 gttgcaggagagatttaa 740
Db 71 GTTGCAGGAGAGATTTTAA 52

RESULT 2
AW229282/c      453 bp mRNA EST 10-DEC-1999
LOCUS      uo08g01.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:2598960 5'
DEFINITION      similar to gb:M33782 TFEF PROTEIN (HUMAN);, mRNA sequence.
ACCESSION      AW229282
VERSION
KEYWORDS
SOURCE
ORGANISM      house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE      1 (bases 1 to 453)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      Unpublished (1997)
JOURNAL
COMMENT      Other ESTs: uo08g01.x1
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

MG1:1022420
Seq primer: -40RP from Gibco
High quality sequence stop: 438.
Location/Qualifiers
1. .453
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="IMAGE:2598960"
/clone_lib="NCL_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Sali;

source
Query Match      1.7%: Score 20; DB 203; Length 458;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 taqtaaaattcaagaattag 1138
Db 250 TAGTAAATTCAGGAATTAG 231

RESULT 4
AI970049      465 bp mRNA EST 25-AUG-1999
LOCUS      wq80d04.x1 NCL_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2477575 3',
DEFINITION

```

```

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      97 a 146 c 148 g 62 t
ORIGIN

Query Match      1.7%: Score 20; DB 112; Length 453;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 taggtctcacagacatactc 37
Db 435 TAGGTCTCAGACACATACTC 416

RESULT 3
AQ165852/c      458 bp DNA GSS 16-OCT-1998
LOCUS      HS_3070_A1.D10.MF CIT Approved Human Genomic Sperm Library D Homo
DEFINITION      sapiens genomic clone Plate=3070 Col=19 Row=G, DNA sequence.
ACCESSION      AQ165852
VERSION      AQ165852.1 GI:3564047
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 458)
AUTHORS      Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL      Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE
COMMENT      Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3070 row: G column: 19
Class: BAC ends
High quality sequence stop: 458.
Location/Qualifiers
1. .458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="plate=3070 Col=19 Row=G"
/sex="male"
/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      93 a 138 c 81 g 144 t 2 others
ORIGIN

Query Match      1.7%: Score 20; DB 203; Length 458;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 taqtaaaattcaagaattag 1138
Db 250 TAGTAAATTCAGGAATTAG 231

RESULT 4
AI970049      465 bp mRNA EST 25-AUG-1999
LOCUS      wq80d04.x1 NCL_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2477575 3',
DEFINITION

```

```

mrna sequence.
ACCESSION AI970049 1 GI:5766867
VERSION AI970049.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 465)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 444.
Location/Qualifiers
1..465
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2477575"
/clone_lib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
985608-986759, 1101192-1101959, and 1217928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 132 a 92 c 101 g 140 t
ORIGIN
Query Match 1.7%; Score 20; DB 104; Length 465;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 729 aagagattttaagagagctg 748
|||||
Db 7 AACAGATTTTAAGAGAGCTG 26

RESULT 5
AA013829/c
LOCUS 604 bp mRNA EST 21-JAN-1997
DEFINITION mh04h05.r1 Soares mouse placenta 4NDMP13.5 14.5 Mus musculus cDNA
clone IMAGE:441561 5' similar to gb:M33782 TFEB PROTEIN (HUMAN);,
mrna sequence.
AA013829
AA013829.1 GI:1474875
EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 604)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:266897
Seq primer: ETPPrimer
High quality sequence stop: 346.
Location/Qualifiers
1..604
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:441561"
/clone_lib="Soares mouse placenta 4NDMP13.5 14.5"
/sex="unknown"
/tissue_type="placenta"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: placenta; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5',
TGTACCAATCTGAAGTGGAGCGCGCGGAATTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 134 a 187 c 166 g 117 t
ORIGIN
Query Match 1.7%; Score 20; DB 1; Length 604;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 18 taggtctcagacagatactc 37
|||||
Db 510 TAGGTCTCAGACAGATAC 491

RESULT 6
BE311177/c
LOCUS 608 bp mRNA EST 26-OCT-2000
DEFINITION BE311177 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3157948 5',
mrna sequence.
BE311177
BE311177.1 GI:9174780
EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 608)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1AM7650 row: c column: 05
 High quality sequence stop: 606.
 Location/Qualifiers

FEATURES

source

```

1. .608
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3157948"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH108"
/notice="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      131 a 173 c 190 g 114 t
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 165; Length 608;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 18 taggtctcagacagatactc 37

|||||

Db 579 TAGGTCTCAGACAGATACTC 560

RESULT 7

```

AI322463/c
LOCUS      AI322463      636 bp      mRNA      EST      23-DEC-1998
DEFINITION      mh04h05.y1 Soares mouse placenta 4NBMP13.5 14.5 Mus musculus cDNA
                  clone IMAGE:441561 5' similar to gb:M33782 TFEB PROTEIN (HUMAN);,
                  mRNA sequence.
ACCESSION      AI322463      GI:4056892
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 636)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and
Waterston,R.

```

```

TITLE      The WashU-HMI Mouse EST Project
JOURNAL
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

```

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:266897

This read is a RESEQUENCE of a previously sequenced mouse clone
 correct orientation)

Seq primer: -40RP from Gibco
 High quality sequence stop: 180.
 Location/Qualifiers

FEATURES

source

```

1. .636
/organism="Mus musculus"

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:441561"
/clone_lib="Soares mouse placenta 4NBMP13.5 14.5"
/sex="unknown"
/tissue_type="placenta"
/lab_host="DH108"
/notice="Organ: placenta; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5',
TGTTCACCAATCTGAAGTGGGCGCGGGAATTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      136 a 203 c 173 g 119 t 5 others
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 18; Length 636;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 18 taggtctcagacagatactc 37

|||||

Db 512 TAGGTCTCAGACAGATACTC 493

RESULT 8

```

AA880933/c
LOCUS      AA880933      665 bp      mRNA      EST      26-MAR-1998
DEFINITION      vx34a05.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
                  IMAGE:1277072 5' similar to gb:M33782 TFEB PROTEIN (HUMAN);, mRNA
                  sequence.
ACCESSION      AA880933      GI:2989916
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 665)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,R. and
Waterston,R.

```

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:668872

Seq primer: -28ml3 revl ET from Amersham

High quality sequence stop: 238.

Location/Qualifiers

```

1. .665
/organism="Mus musculus"
/strain="C57BL/6 x CBA"
/db_xref="taxon:10090"
/clone="IMAGE:1277072"
/clone_lib="Stratagene mouse lung 937302"
/sex="female"
/tissue_type="lung"

```



```

/dev_stage="6-8 month old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. 6-8 month old female lung and 1.5 year old male lung
were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP
XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3'
adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT      149 a   203 c   191 g   122 t
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 13; Length 665;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 taggtctcagacagatactc 37
|||||
Db 318 TAGGTCTCAGACAGATCTC 299

```

```

RESULT 9
AZ693348/c      842 bp      DNA      GSS      14-DEC-2000
LOCUS      ENTML05TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION      genomic, DNA sequence.
ACCESSION      AZ693348
VERSION      AZ693348.1 GI:11830494
KEYWORDS      GSS.
SOURCE      Entamoeba histolytica.
ORGANISM      Eukaryota; Entamoebidae; Entamoeba.

```

```

REFERENCE
AUTHORS      Loftus.B., Van Aken.S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
JOURNAL      HMI:IMSS sheared DNA library
COMMENT      Unpublished (2000)
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: bjloftus@tigr.org
              Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
              DNA library
              Seq primer: M13-Forward
              Class: shotgun
              High quality sequence start: 32
              High quality sequence stop: 676.
              Location/Qualifiers
                1..842
                  /organism="Entamoeba histolytica"
                  /strain="HMI:IMSS"
                  /db_xref="taxon:5759"
                  /clone_lib="Entamoeba histolytica Sheared DNA"
                  /note="Vector: pHOSt1; Site_1: Bst I; Constructed at The
                  Institute for Genomic Research (TIGR), Rockville, MD.
                  Genomic DNA isolated from broth cultures of E. histolytica
                  using a method described by Clark and Diamond (Clark,
                  C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                  method for isolate identification. Exp. Parasitol.
                  77:450.). The DNA was mechanically sheared to give a
                  tight size distribution (~2 kb). The v + i method used for
                  the library construction is described in detail in Smith,
                  H.O. and Venter, J.C. (Making small insert libraries for
                  whole genome shotgun sequencing projects. In Genome
                  Sequencing: A Practical Approach, eds. M. Vaudin and B.
                  Borell, Oxford University Press, 1999)."

```

```

BASE COUNT      300 a   82 c   77 g   383 t
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 226; Length 842;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 774 attatatactaaacataata 793
|||||
Db 171 ATTATATACTAAACATAATA 152

```

```

RESULT 10
AZ675337/c      893 bp      DNA      GSS      14-DEC-2000
LOCUS      ENTW55TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION      genomic, DNA sequence.
ACCESSION      AZ675337
VERSION      AZ675337.1 GI:11812603
KEYWORDS      GSS.
SOURCE      Entamoeba histolytica.
ORGANISM      Eukaryota; Entamoebidae; Entamoeba.

```

```

REFERENCE
AUTHORS      Loftus.B., Van Aken.S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
JOURNAL      HMI:IMSS sheared DNA library
COMMENT      Unpublished (2000)
              Contact: Brendan J Loftus
              Department of Eukaryotic Genomics
              The Institute for Genomic Research
              9712 Medical Center Dr., Rockville, MD 20850, USA
              Tel: 301 838 0200
              Fax: 301 838 0208
              Email: bjloftus@tigr.org
              Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
              DNA library
              Seq primer: M13-Reverse
              Class: shotgun
              High quality sequence start: 20
              High quality sequence stop: 873.
              Location/Qualifiers
                1..893
                  /organism="Entamoeba histolytica"
                  /strain="HMI:IMSS"
                  /db_xref="taxon:5759"
                  /clone_lib="Entamoeba histolytica Sheared DNA"
                  /note="Vector: pHOSt1; Site_1: Bst I; Constructed at The
                  Institute for Genomic Research (TIGR), Rockville, MD.
                  Genomic DNA isolated from broth cultures of E. histolytica
                  using a method described by Clark and Diamond (Clark,
                  C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                  method for isolate identification. Exp. Parasitol.
                  77:450.). The DNA was mechanically sheared to give a
                  tight size distribution (~2 kb). The v + i method used for
                  the library construction is described in detail in Smith,
                  H.O. and Venter, J.C. (Making small insert libraries for
                  whole genome shotgun sequencing projects. In Genome
                  Sequencing: A Practical Approach, eds. M. Vaudin and B.
                  Borell, Oxford University Press, 1999)."

```

```

BASE COUNT      356 a   106 c   91 g   340 t
ORIGIN

```

```

Query Match      1.7%; Score 20; DB 225; Length 893;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 774 attatatactaaacataata 793
|||||
Db 801 ATTATATACTAAACATAATA 782

```

```

RESULT 11
AA492011/c      175 bp      mRNA      EST      19-AUG-1997
LOCUS

```

```

DEFINITION ne72h01.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:909841, mRNA
sequence.
ACCESSION AA492011
VERSION AA492011.1 GI:2221573
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 305 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES
source
1..175
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:909841"
/clone="NCI_CGAP_Ew1"
/tissue_type="Ewing's sarcoma"
/lab_host="DH10B"
/Note="Vector: pAMP10; mRNA made from Ewing's sarcoma,
cDNA made by oligo-dT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 53 a 44 c 29 g 49 t
ORIGIN
Query Match 1.6%; Score 19; DB 8; Length 175;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1086 acgtataaatgataatca 1104
|||||
Db 170 ACGTAAATGATAATCA 152

RESULT 12
BF362820 272 bp mRNA EST 24-NOV-2000
LOCUS CM4-NN0079-050600-191-a05 NN0079 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF362820
ACCESSION BF362820.1 GI:11324845
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 272)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

```

```

JOURNAL Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics.
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM4&t2=CM4-NN0079-
050600-191-a05&t3=2000-06-05&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 272.
FEATURES
source
1..272
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0079"
/dev_stage="Adult"
/Note="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 105 a 38 c 39 g 90 t
ORIGIN
Query Match 1.6%; Score 19; DB 147; Length 272;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 726 aggaagagatttaagaga 744
|||||
Db 171 AGGAAGAGATTTTAAAGAGA 189

RESULT 13
AV136611/c
LOCUS AV136611 291 bp mRNA EST 01-JUL-1999
DEFINITION AV136611 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
clone 2810027G08, mRNA sequence.
ACCESSION AV136611
VERSION AV136611.1 GI:5322341
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 291)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara
, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara
, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomimaga, N.,
Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@res.riken.go.jp
Thermotabilization and thermoactivation of thermolabile enzymes by

```

trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

1. .291
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2810027608"
/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"
70 a 67 c 47 g 107 t

BASE COUNT

ORIGIN

Query Match 1.6%; Score 19; DB 109; Length 291;

Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 352 gaagggttaaaactgagg 370

|||||

Db 120 GAAGGGTTAAAACTGAGG 102

RESULT 14

H18381/c

LOCUS

H18381 331 bp mRNA EST 29-JUN-1995

yn4901.r1 Soares adult brain N2b5HB55y Homo sapiens cDNA clone

IMAGE:171792 5', mRNA sequence.

ACCESSION

H18381

VERSION

H18381.1 GI:884621

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 331)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman

, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,

Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston

, R., Williamson, A., Wohldmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 705

High quality sequence stops: 268

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 705 Std Error: 0.00

Seq primer: M13RPI

High quality sequence stop: 268.

Location/Qualifiers

1. .331

/organism="Homo sapiens"

/db_xref="GDB:3833827"

/db_xref="taxon:9606"

/clone="IMAGE:171792"

/clone_lib="Soares adult brain N2b5HB55y"

/sex="Male"

/dev_stage="55-year old"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: brain; Vector: p77f3D (Pharmacia) with a

modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCAATCTGAAGTGGAGCGCGCTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified p77f3 vector (Pharmacia). Library went through one round of normalization to a Cot = 53. Library constructed by Bento Soares and M. Fatima Bonaldo. The adult brain RNA was provided by Dr. Donald H. Gilden. Tissue was acquired 17-18 hours after death which occurred in consequence of a ruptured aortic aneurysm. RNA was prepared from a pool of tissues representing the following areas of the brain: frontal, parietal, temporal and occipital cortex from the left and right hemispheres, subcortical white matter, basal ganglia, thalamus, cerebellum, midbrain, pons and medulla."

BASE COUNT

ORIGIN

104 a 52 c 62 g 111 t 2 others

Query Match 1.6%; Score 19; DB 152; Length 331;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1110 tgaaaaagctagtaaaatt 1128

|||||

Db 228 TGAAAAAGCTAGTAAATTT 210

RESULT 15

AI638554

LOCUS

AI638554

ts50a01.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2231976 3',

mRNA sequence.

AI638554

AI638554.1 GI:4690788

EST.

KEYWORDS

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 365)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Sequencing by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 2276 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 360.

Location/Qualifiers

1. .365

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2231976"

/clone_lib="NCI_CGAP_Ut1"

/tissue_type="well-differentiated endometrial

adenocarcinoma, 7 pooled tumors"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

FEATURES

source

BASE COUNT 141 a 58 c 54 g 112 t
ORIGIN

Query Match 1.6% Score 19; DB 23; Length 365;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1110 tgaataagctagtaaaatt 1128
 |||||
DB 345 TGAAAAAGCTAGTAAATTT 363

Search completed: May 13, 2001, 05:47:55
Job time: 3135 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 01:59:25 ; Search time 1527.71 Seconds
(without alignments)
11178.837 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagcatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_bal.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_bal.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_vi.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_vil.*
59: gb_vil2.*
60: gb_vil3.*
61: gb_vil4.*
62: gb_vil5.*
63: gb_vil6.*
64: gb_vil7.*
65: gb_vil8.*
66: gb_vil9.*
67: gb_vil10.*
68: gb_vil11.*
69: gb_vil12.*
70: gb_vil13.*
71: gb_vil14.*
72: gb_vil15.*
73: gb_vil16.*
74: gb_vil17.*
75: gb_vil18.*
76: gb_vil19.*
77: gb_vil20.*
78: gb_vil21.*
79: gb_vil22.*
80: gb_vil23.*
81: gb_vil24.*
82: gb_vil25.*
83: gb_vil26.*
84: gb_vil27.*
85: gb_vil28.*
86: gb_vil29.*
87: gb_vil30.*
88: gb_vil31.*
89: gb_vil32.*
90: gb_vil33.*
91: gb_vil34.*
92: gb_vil35.*
93: gb_vil36.*
94: gb_vil37.*
95: gb_vil38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	419	36.2	741	3	S64396
2	241.6	20.9	2131	3	STREMM24
3	175	15.1	1547	3	STRSMP3A
4	171.8	14.8	2633	3	SPU02480
5	158	13.6	1817	2	AF288752
6	158	13.6	1620	3	SPU40231
7	156.4	13.5	1465	3	STREMM3
8	154.8	13.4	1746	3	SPENM3G
9	151	13.0	229	3	S79254
10	151	13.0	237	3	SPU11940
11	151	13.0	300	3	SPU20094

12	151	13.0	1332	9	AR079002	AR079002 Sequence
13	151	13.0	1494	3	SPENM1M	X62131 S.pyogenes
14	151	13.0	1708	3	SPENM1	X07933 S.pyogenes
15	149.4	12.9	251	3	SPU11997	U11997 Streptococc
16	149.4	12.9	300	3	SPU20097	U20097 Streptococc
17	149.4	12.9	300	3	SPU20099	U20099 Streptococc
18	149.4	12.9	300	3	SPU20100	U20100 Streptococc
19	147.4	12.7	907	1	AF060760	AF060760 Streptoco
20	137	11.8	694	2	AF288753	AF288753 Streptoco
21	130	11.2	249	3	SPAJ3285	AJ223285 Streptoco
22	123.8	10.7	300	3	SPU20095	U20095 Streptococc
23	123.8	10.7	1455	3	SPM11PROA	221845 S.pyogenes
24	113	9.8	261	3	SPU11945	U11945 Streptococc
25	108.8	9.4	296	3	SPU11959	U11959 Streptococc
26	106.4	9.2	451	3	SPU11986	U11986 Streptococc
27	106.4	9.2	2111	3	STRM6	M11338 Streptococc
28	105.6	9.1	301	10	I01016	I01016 Sequence 1
29	104.8	9.1	249	3	SPAJ3286	AJ223286 Streptoco
30	104.6	9.0	300	3	SPU20098	U20098 Streptococc
31	97.4	8.4	310	3	SPU20058	S82058 N-terminal
32	97.4	8.4	311	3	SPENMT19	X56609 S.pyogenes
33	92.8	8.0	300	3	SPU20096	U20096 Streptococc
34	92.4	8.0	549	2	AF281048	AF281048 Streptoco
35	92	7.9	664	2	AF222860	AF222860 Streptoco
36	91.6	7.9	321	3	SGSM507	X79527 Streptococc
37	91.4	7.9	1677	3	SPU39838	U39838 Streptococc
38	90.6	7.8	263	3	SPU11943	U11943 Streptococc
39	89.4	7.7	299	3	STREMM32G	L47325 Streptococc
40	85.2	7.4	225	1	AF018179	AF018179 Streptoco
41	84.6	7.3	430	2	G025741	U25741 Group G str
42	84.6	7.3	450	3	SGSM935	X79533 Streptococc
43	84	7.3	261	3	STREMM	L46652 Streptococc
44	82.6	7.1	7218	10	I66494	I66494 Sequence 14
45	77.6	6.7	714	2	AF325839	AF325839 Streptoco

RESULT	1					
LOCUS	S64396	741 bp	DNA	BCT	15-OCT-1993	
DEFINITION	emm=tetravalent M protein [Streptococcus, group A, Genomic Recombinant, 741 nt].					
ACCESSION	S64396					
VERSION	S64396.1	GI:408224				
KEYWORDS	Streptococcus group A.					
SOURCE	Streptococcus					
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae.					
REFERENCE	1 (bases 1 to 741)					
AUTHORS	Dale, J.B., Chiang, E.Y. and Lederer, J.W.					
TITLE	Recombinant tetravalent group A streptococcal M protein vaccine					
JOURNAL	J. Immunol. 151 (4), 2188-2194 (1993)					
MEDLINE	93346765					
REMARK	GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 136070] from the original journal article. This sequence comes from Fig. 1.					
FEATURES	Location/Qualifiers					
source	1..741					
	/organism="Streptococcus"					
	/db_xref="taxon:1301"					
gene	1..741					
	/partial					
	/gene="emm"					
	/note="tetravalent M protein"					
CDS	1..741					
	/partial					
	/gene="emm"					
	/note="Hybrid molecule containing amino-terminal subunits of types 24, 5, 6, and 19 M proteins; This sequence comes from Fig. 1"					
	/codon_start=1					

1	36.2%;	Score 419;	DB 3;	Length 741;
Best Local Similarity	78.9%;	Pred. No. 1.9e-67;		
Matches	585;	Conservative	0;	Mismatches
			156;	Gaps
			1;	

QY	7	atggtcggaactaggtctcagacagatactctgaaagaaagtaacaaagcgtgctgacaag	66
Db	1	ATGGTTCGGCGACTAGTCTCAGACAGACTCTCTGAAAAAGTACAAAGACCTGCTGACAAG	60
QY	67	tgtgagatagaaaaacaatcacgttaaaacttaagaatagtgacttaagctttaataataaa	126
Db	61	TTTGAGATAGAAAAACAATACGTTAAACTTAAGAAATAGTACTTAAGTTTAATAATAAA	120
QY	127	gcgttaagaatcataagtagagtgagttactaaagagtgagtaagctaaagagaacta	186
Db	121	GGGTTAAAGATCATATATGATGAGTTAACTGAAGAGTTGAGTAATGCTTAAGAGAACTA	180
QY	187	cgt-----	189
Db	181	CGTAAAAATGATAAATCACTATCTCAAAAAAGCTAGTAAAAATCAAGAATTAGAGGCACGT	240
QY	190	-----	189
Db	241	AAGGCTGATCTTGAANAACATTAGAGCGCAATGAATTTTCACAGCGGATTCAGCT	300
QY	190	-----	210
Db	301	AAATCAAAACCTTAGACGAGAGAAAGCTGATCTTGAAGGATCGCGGTGACTAGGGGT	360
QY	211	acaataatgacccgcaagagagcaaaagagctcttgacaagctatgagctagaaacacat	270
Db	361	ACAATAATGACCCGCAAGAGAGCAAAAGAGCTCTTTGACCAAGTATGAGCTAGAAAACCAT	420
QY	271	gacttaaaactaagaatgaagggttaaaaactgagaatgaagggttaaaaactgagaat	330
Db	421	GACTTAAAACTAAGATGAAGGTTAAAAACTGAGAATGAAGGTTAAAAACTGAGAAT	480
QY	331	gaagggttaaaaactgagaatgaagggttaaaaactgagaatgagctgacagagctgttctagg	390
Db	481	GAGGGTTAAAACTGAGAATGAAGGTTAAAAACTGAGGTGACAGAGTGTTCCTAGG	540
QY	391	ggagcgtagaaaaacccggagacacacgacagaaacttcttaacaagtatgacgtagagaac	450
Db	541	GGGACGGTAGAAAAACCCGGCAAGACGACGAGAACTCTTTAAACAAGTATGACGTAGAGAAC	600
QY	451	tctatgttacaagctaataatgacaagtaccatgagagagtgagttactactagacatac	510
Db	601	TCTATGTTACAAAGCTAATAATGACAAGTTACCATGGAGAGTGGCTTATATAGGCATACG	660
QY	511	ccagaagataagctaaaaaaattattgacgactcttgacgcaaaagaaacatgaattacaa	570
Db	661	CCAGAAGATAGCTTAAAAAATAATTTATGACGNTCTTCGCGCAAAAGAACATGAATTACAA	720
QY	571	caacgaatgagaagttatct	591
Db	721	CAACGAATGAGAACTTATCT	741

RESULT	2					
STREMM24	2131 bp	DNA	BCT	26-APR-1993		
LOCUS	STREMM24					
DEFINITION	S.pyogenes type 24 M-protein gene, complete cds.					

LOCUS STRSMF5A 1547 bp DNA ECT 26-APR-1993
DEFINITION S.Pyogenes smf5 gene encoding serotype 5 M protein, complete cds.
ACCESSION M20374
VERSION M20374.1 GI:153812
KEYWORDS 5 M protein; smf5 gene;
SOURCE S.pyogenes (strain Manfredo) DNA, clone pmk207.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.

REFERENCE 1 (bases 1 to 1547)
AUTHORS Miller, L., Gray, L., Beachey, E.H. and Kehoe, M.
TITLE Antigenic variation among group A streptococcal M proteins:
Nucleotide sequence of the serotype 5 M protein gene and its
relationship with genes encoding types 6 and 24 M proteins
J Biol Chem. 263, 5668-5673 (1988)

JOURNAL 88186881 Location/Qualifiers
FEATURES
source 1..1547
organism="Streptococcus pyogenes"
db_xref="taxon:1314"
43..1521
note="5 M protein"
codon_start=1
transl_table=1
protein_id="AAA26976.1"
db_xref="GI:153812"
translation="MARENTKHYWLKLKKGTASVAVALSVLGAGLVVNTNEVSAAV
TRGTVNDPQAKALKDQYLENHDLTKNEGDKTENEGKTENGLKTEKGILTKPKK
EHEANDK LQQRDVTSTOKETLEREVQNTQYNNETLKINGDUTLKLNTKROELANK
QDSKEKNALNELKETVDKRIAEQEENKETIGTLKIHDETVDKRIAEQEENKETKI
GLKKIHDETVKLDLAKRKQNKONIGALKQLAKELKADKANIKSDASKRGRLRRDDLSGR
AKLOEDAHKOLEEQNKI SEASRKGRLRDDLSREARKQIEAQOOLKEQNKI SEASR
KGLRRDLDAKQKVKVEKALFEANSKLAALKLNKELESSEK KL TEKEKAELQAK LEA
EAKALKPOLAKQAQEKVLKRAGKASDSQTPTDPGNKAVPGKGQAPQAGT KPNQNAK
MKETKRLPSTGETNPFPFTAAALTVMATAGVAAVVVRKKEEN"

	repeat_region	238..258	/note="repeat type A"
repeat_region	259..279	/note="repeat type A"	
repeat_region	280..300	/note="repeat type A"	
repeat_region	301..321	/note="repeat type A"	
repeat_region	322..342	/note="repeat type A"	
repeat_region	343..363	/note="repeat type A"	
repeat_region	364..384	/note="repeat type A"	
repeat_region	385..405	/note="repeat type B"	
repeat_region	406..426	/note="repeat type B"	
repeat_region	427..447	/note="repeat type B"	
repeat_region	448..468	/note="repeat type B"	
repeat_region	469..489	/note="repeat type C"	
repeat_region	490..510	/note="repeat type C"	
repeat_region	511..531	/note="repeat type C"	
repeat_region	532..552	/note="repeat type C"	
repeat_region	553..573	/note="repeat type C"	
repeat_region	574..594	/note="repeat type C"	
repeat_region	595..615	/note="repeat type C"	
repeat_region	616..636	/note="repeat type C"	
repeat_region	637..657	/note="repeat type C"	
repeat_region	658..678	/note="repeat type C"	
repeat_region	679..699	/note="repeat type C"	
repeat_region	700..720	/note="repeat type C"	
repeat_region	721..741	/note="repeat type C"	
repeat_region	742..762	/note="repeat type C"	
repeat_region	763..783	/note="repeat type C"	
repeat_region	784..804	/note="repeat type C"	
repeat_region	805..825	/note="repeat type C"	
repeat_region	826..846	/note="repeat type C"	
repeat_region	847..867	/note="repeat type C"	
repeat_region	868..888	/note="repeat type C"	
repeat_region	889..909	/note="repeat type C"	
repeat_region	910..930	/note="repeat type C"	
repeat_region	931..951	/note="repeat type C"	
repeat_region	952..972	/note="repeat type C"	
repeat_region	973..993	/note="repeat type C"	
repeat_region	994..1014	/note="repeat type C"	
repeat_region	1015..1035	/note="repeat type C"	
repeat_region	1036..1056	/note="repeat type C"	
repeat_region	1057..1077	/note="repeat type C"	
repeat_region	1078..1098	/note="repeat type C"	
repeat_region	1099..1119	/note="repeat type C"	
repeat_region	1120..1140	/note="repeat type C"	
repeat_region	1141..1161	/note="repeat type C"	
repeat_region	1162..1182	/note="repeat type C"	
repeat_region	1183..1203	/note="repeat type C"	
repeat_region	1204..1224	/note="repeat type C"	
repeat_region	1225..1245	/note="repeat type C"	
repeat_region	1246..1266	/note="repeat type C"	
repeat_region	1267..1287	/note="repeat type C"	
repeat_region	1288..1308	/note="repeat type C"	
repeat_region	1309..1329	/note="repeat type C"	
repeat_region	1330..1350	/note="repeat type C"	
repeat_region	1351..1371	/note="repeat type C"	
repeat_region	1372..1392	/note="repeat type C"	
repeat_region	1393..1413	/note="repeat type C"	
repeat_region	1414..1434	/note="repeat type C"	
repeat_region	1435..1455	/note="repeat type C"	
repeat_region	1456..1476	/note="repeat type C"	
repeat_region	1477..1497	/note="repeat type C"	
repeat_region	1498..1518	/note="repeat type C"	
repeat_region	1519..1539	/note="repeat type C"	
repeat_region	1540..1560	/note="repeat type C"	
repeat_region	1561..1581	/note="repeat type C"	
repeat_region	1582..1602	/note="repeat type C"	
repeat_region	1603..1623	/note="repeat type C"	
repeat_region	1624..1644	/note="repeat type C"	
repeat_region	1645..1665	/note="repeat type C"	
repeat_region	1666..1686	/note="repeat type C"	
repeat_region	1687..1707	/note="repeat type C"	
repeat_region	1708..1728	/note="repeat type C"	
repeat_region	1729..1749	/note="repeat type C"	
repeat_region	1750..1770	/note="repeat type C"	
repeat_region	1771..1791	/note="repeat type C"	
repeat_region	1792..1812	/note="repeat type C"	
repeat_region	1813..1833	/note="repeat type C"	
repeat_region	1834..1854	/note="repeat type C"	
repeat_region	1855..1875	/note="repeat type C"	
repeat_region	1876..1896	/note="repeat type C"	
repeat_region	1897..1917	/note="repeat type C"	
repeat_region	1918..1938	/note="repeat type C"	
repeat_region	1939..1959	/note="repeat type C"	
repeat_region	1960..1980	/note="repeat type C"	
repeat_region	1		

BASE COUNT		1153 a		463 c		532 g		485 t	
ORIGIN									
<pre> /label=Prot-enn5 /product="Enn5_8193 protein" /protein_id="AAA50855.1" /db_xref="GI:437192" /translation="MTROQTKNYSLRKLTGTASVAVALTVLGAFNQTVEYKAEGY RACSTIANSKISLYDENSKLVEERAELIDKIDKLEELAKEKKAEDKAEDVERHY LRKLOEHEKEEROKNLEELEROGKREIDKRYOELOKQOQOQLETKELTEANQVSE LRKGLSNDLEASRAKKELEAKHOKLEADHOVSETSRKGLSRDLEASREANKKVTSE LTKQAQUSALEESKKLTEKEKAELOAKLEAKFKEOLAKQAEELAKLERAGKVSDS QTPEADKPGNKVYPGKGQAPQAGAPNQNKAPMKETKQLPSTGETANPPTFAAALTVM ATAGVAAYVYKKEEN" </pre>									
Query Match		14.8%		Score 171.8		DB 3		Length 2633	
Best Local Similarity		91.5%		Pred. No. 3.4e-22		O: 0			
Matches 182		Conservative		0		Mismatches 17		Indels 0	
		Gaps		0					
QY	171	tgctaaagagaacacgctggtgacgcgcgctgactaggggtacaataataatgacccgcaaaag	230						
DB	30	TGTCATACTAATCAAGTTAGTGCACCGCTGACTAGGGGTACAATAAATGACCCGCAAAAG	89						
QY	231	agcaaaagaactcttgacaagttagctgactagaaaaccatgacttaaaaaactaagaatga	290						
DB	90	AGCAAAAGAAGCTCTTGACAAGTATGAGCTAGAAAACCATGACTTAAAAACTAAGAATGA	149						
QY	291	agggttaaaaactgagaatgaaggggttaaaaactgagaatgaaggggttaaaaactgagaa	350						
DB	150	AGGGTTAAAAACTGAGATGAAGGGTTAAAAACTGAGAATCAAGGGTTAAAAACTGAGAA	209						
QY	351	tgaagggttaaaaactgag	369						
DB	210	TGAAGGTTAAAAAACTGAG	228						
RESULT	5								
AF288752	AF288752 817 bp DNA BCT 02-AUG-2000								
LOCUS	Streptococcus pyogenes isolate 3148-00 M protein (emm) gene,								
DEFINITION	partial cds.								
ACCESSION	AF288752								
VERSION	AF288752.1 GI:9652381								
KEYWORDS									
SOURCE	Streptococcus pyogenes.								
ORGANISM	Streptococcus pyogenes								
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae								
AUTHORS	Beall,B.W.								
TITLE	5' emm sequence emm3.2								
JOURNAL	Unpublished								
REFERENCE	2 (bases 1 to 817)								
AUTHORS	Beall,B.W.								
TITLE	Direct Submission								
JOURNAL	Submitted (21-JUL-2000) Respiratory diseases Branch, Centers for								
	Disease Control and Prevention, 1600 Clifton Rd. NE, Atlanta, GA								
	30333, USA								
FEATURES	Location/Qualifiers								
source	1. .817								
	/organism="Streptococcus pyogenes"								
	/isolate="3148-00"								
	/db_xref="taxon:1314"								
	/note="blood isolate recovered from patient in New York								
	population based surveillance; T non typeable; opacity								
	factor negative								
	group: A"								
	<1. .>817								
gene	/gene="emm"								
	<1. .>817								
CDS	/gene="emm"								
	/note="emm3.2"								
	/codon_start=1								

BASE COUNT 342 a 105 c 182 g 188 t
ORIGIN

/transl_table=11
/product="M protein"
/protein_id="AAF91496.1"
/db_xref="GI:9652382"
/translation="TASVAVALTVLTGLVAGOTVKADARSVNGFPRHVKLN
EINLLDPVLTQYKINSNYQYNAQAGRLDROKAYLKLGLNDWAERLLOELNGEDV
KKVLGVAFKEDDLKEVKEKIDKKEKEODLDKDFLAKGCVLSDRHQEULE
EKEKVTETAKVGOISSELETVKQVSEETVQKVESTMOQDTEKQNRYSOLSEELATPKQAKED
FELAAANAADKOKLEAKIADLETKLEAKEDFELALGHORAHNEYQKLAEK"

Query Match 13.6%; Score 158; DB 2; Length 817;
Best Local Similarity 63.4%; Pred. No. 1.2e-19;
Matches 242; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 747 tgggtacctgttagatcaggttacacaattatatactatacaataatagtaattaccaca 806
Db 138 TGAGAACTGTTAGATCAGGTACACAATTATATCTAAACATAATAGTATTACCAACA 197

Qy 807 atataatgcacagctgcagacttgacctgacagacacaaagctgaatatctaaagcct 866
Db 198 ATATATGCACAGCTGCGACACTTGACCTGACACAAAGGCTGATATCTTAAGGCGCT 257

Qy 867 taatgattg99gtgagggctgtgtacagagttataatctcgatgcgcgactaggtctca 926
Db 258 TAATGATGTTGGCTGAGAGCGTGTACAAAGAGTGTAAATGGAGAGATGTAAAAAAGT 317

Qy 927 gacagatctctggaaaaaglacagaacgctgctgacaaagtttgagatgagaaaacaatc 986
Db 318 AGGTAAGTGGCTTTTGAAGAAAGATGATTTAGAAAAGGAGGTAAAGAACTTAAGGAAA 377

Qy 987 gttaaaacttaagaatgacttaagtttttaataataaagctgtaaaagctcatatga 1046
Db 378 AATAGACAAAGGAAGGAATATCAGGACTTAGATTAAGATTTTGACTTGGCCAAACA 437

Qy 1047 tgaagttaactgagagtgagtaatgctaaagagaaactacataaaatgataaatacct 1106
Db 438 GGGGTATGTTTATCAGATTAAGACATCAACAGAACTAGAGAAGAAAGAAAGAAAGT 497

Qy 1107 atctgaaaaagctagtaaaatt 1128
Db 498 TACAGAAGCAACTGCTAAAGTT 519

RESULT 6
SPU40231 1620 bp DNA BCT 15-APR-1996
LOCUS Streptococcus pyogenes M3 protein (emm3.1) gene, complete cds.
DEFINITION
ACCESSION U40231
VERSION U40231.1 GI:1263022
KEYWORDS
SOURCE Streptococcus pyogenes strain=Type 3.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 1620)
AUTHORS Dale,J.B., Simmons,M., Chiang,E. and Chiang,E.
TITLE Octavalent group A streptococcal M protein vaccine
JOURNAL Vaccine (1996) In press
REFERENCE 2 (bases 1 to 1620)
AUTHORS Dale,J.B.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1995) James B. Dale, Medicine, VA Medical Center,
1030 Jefferson Ave., Memphis, TN 38104, USA
FEATURES
source location/Qualifiers
1. .1620
/organism="Streptococcus pyogenes"
/strain="type 3"
/db_xref="taxon:1314"
1. .1620
/gene="emm3.1"

CDS 1. .1620
/gene="emm3.1"
/codon_start=1
/transl_table=11
/product="M3 protein"
/protein_id="AAA96960.1"
/db_xref="GI:1263023"
/translation="MAKNNTNRHYSRLKTKTASVAVLTVLTGLVAGOTVKADAR
SVNGFPRHVKLNKEINLLDPVLTQYKINSNYQYNAQAGRLDROKAYLKLGLNDWAERLLOELNGEDV
KKVLGVAFKEDDLKEVKEKIDKKEKEODLDKDFLAKGCVLSDRHQEULEKEDFELALGHORAH
NEYQKLEAKEDFELAAANAADKOKLEAKIADLETKLEAKEDFELALGHORAH
VTEOKQIILDASRKGTARDLEAVROAKQOEALQLEENRISSEKSGKRRDLDSAR
EAKQVKEALEEANSKLAALKELEESKLEKEKLEAKLEAKLEAKLEAKLEAKLEAKLEAKLEAK
QAELEAKLRAGKASDSQIPDTPKGNKAVPKGQAPQAGTKPNONKAPMKETKRLPST
GETANPFTAAALTVMATAGVAAVVRKEEN"

BASE COUNT 697 a 269 c 335 g 319 t
ORIGIN

Query Match 13.6%; Score 158; DB 3; Length 1620;
Best Local Similarity 63.4%; Pred. No. 1.2e-19;
Matches 242; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 747 tgggtacctgttagatcaggttacacaattatatactatacaataatagtaattaccaca 806
Db 180 TGAGAACTGTTAGATCAGGTACACAATTATATCTAAACATAATAGTATTACCAACA 239

Qy 807 atataatgcacagctgcagacttgacctgacagacacaaagctgaatatctaaagcct 866
Db 240 ATATAATGCACAGCTGCGACACTTGACCTGAGACAAAAGCTGAATATCTAAAGGCT 299

Qy 867 taatgattg99gtgagggctgtgtacagagttataatctcgatgcgcgactaggtctca 926
Db 300 TAATGATGTTGGCTGAGAGCGTGTACAGAGTTAAATGGAGAGATGTAAAGAAAGT 359

Qy 927 gacagatctctggaaaaaglacagaacgctgctgacaaagtttgagatgagaaaacaatc 986
Db 360 AGGTAAAGTGGCTTTTGAAGAAAGATGATTTAGAAAAGGAGGTAAAGAACTTAAGGAAA 419

Qy 987 gttaaaacttaagaatgacttaagtttttaataataaagctgtaaaagctcatatga 1046
Db 420 AATAGACAAAGGAAGGAATATCAGGACTTAGATTAAGATTTTGACTTGGCCAAACA 479

Qy 1047 tgaagttaactgagagtgagtaatgctaaagagaaactacataaaatgataaatacct 1106
Db 480 GGGGTATGTTTATCAGATTAAGACATCAACAGAACTAGAGAAGAAAGAAAGAAAGT 539

Qy 1107 atctgaaaaagctagtaaaatt 1128
Db 540 TACAGAAGCAACTGCTAAAGTT 561

RESULT 7
STREMM3 1465 bp DNA BCT 01-FEB-2000
LOCUS S. pyogenes DNA for M 3 protein, partial sequence.
DEFINITION
ACCESSION D14415
VERSION D14415.1 GI:511939
KEYWORDS M 3 protein.
SOURCE Streptococcus pyogenes (strain:C203 (ATCC 12384)) DNA.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 1465)
AUTHORS Katsukawa,C.
TITLE Cloning and nucleotide sequence of the type 3 M protein gene (emm3)
consisting of an N-terminal variable portion and C-terminal
conserved C-repeat regions: Relation to other genes of
Streptococcus pyogenes
J. Jpn. Assoc. Infect. Dis. 68, 698-705 (1994)
REFERENCE 2 (bases 1 to 1465)


```
/db_xref="SPTREMBL:Q54835"
/translation="MAKNNTNRHYSRLRKLKGTASVAVALTVLGTGLVAGQTVKADAR
SVNGEPRHVKLENEJENLDQVOTLYNKHNSNTQOYSAQAGRLDLRQRAEYKGLND
WAERLLQELNGEDVKVLFKEDDKLEKEKIDKKEKIDKKEKIDKKEKIDKKEKIDKKE
GYVLSDRHQHQLKEKEKIDKKEKIDKKEKIDKKEKIDKKEKIDKKEKIDKKEKIDKKE
BQELATTQNAKEDFELAAANADKQLEAKIADLETKLKEAKEDFELAAALGHQHAH
NEYQAKLEAKEDFELAAALADKQLEAKIADLETKLKEAKEDFELAAALGHQHAH
VTEQKQILDASRKGTARDLEAVRQAKQLEAKIADLETKLKEAKEDFELAAALGHQHAH
BAKQVQKDLANTLELDKQLEAKIADLETKLKEAKEDFELAAALGHQHAH
LALEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE
QIPDTGKNAVPGKGOAQAQCTKPNQKAPKKTETKQLPSTGETANPPTAAALTVM
ATAGVAAVVKEEN"
1..1746
/gene="emm3"
/mat_peptide
124..1743
/gene="emm3"
/product="M3 protein"
BASE COUNT 742 a 292 c 368 g 344 t
ORIGIN
Query Match 13.4%; Score 154.8; DB 3; Length 1746;
Best Local Similarity 62.8%; Pred. No. 4.4e-19;
Matches 240; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
QY 747 tggtagcttggtagctgggtacacattatatactaaacataatagtaattaccaca 806
DB 180 TGAGAACTTGTAGATCAGGTACACAAATATATAATAACATATAATATACCAACA 239
QY 807 ataatgcacagctggcagacttgacctgagacaaagggctgtaattctaaaggcct 866
DB 240 ATATAGTGCACAGCTGGCAGACTTGACCTGACACAAAGAGGCTGAATATCTAAAAGGCT 299
QY 867 taatgattggctgagagctgttacaagagttaaatatcgtcgcgactaggtctca 926
DB 300 TAATGATTGGCTGAGAGCTGTTACAGAGTTAAATGCGAGAGATGTAAAAAGTTT 359
QY 927 gacagactctggaaaagtacaagacgtgctgcaagatttgagatagaaaacaatac 986
DB 360 AGGTAAAGTGGCTTTGAAAGAGATGATTTAGAAAGGAGGTTAAAGAACTTAAGGAAA 419
QY 987 gtaaaacttaagaatagtgacttaagttttaataataaagcgttaaaagatcataatga 1046
DB 420 AATAGACAAAAGGAAAGGAATATCAGGACTTAGATAAAGATTTTGACTTGGCCAAACA 479
QY 1047 tgagttaactgaagattgagtaagtctaaagagaaactacgttaaaatgataaatacact 1106
DB 480 GGGGTATGTTTATCATGATAAAGACATCAACAAGAACTAGAGAAGAAAGAAAGAAAGT 539
QY 1107 atctgaaaagctagttaaaatt 1128
DB 540 TACAGAGCACTGCTAAGTT 561
RESULT 9
S79254
LOCUS S79254 229 bp DNA BCT 30-NOV-1995
DEFINITION emm-M protein [Streptococcus pyogenes, M1a, Genomic, 229 nt].
ACCESSION S79254
VERSION S79254.1 GI:1087104
KEYWORDS
SOURCE Streptococcus pyogenes M1a.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 229)
Penney,T.J., Martin,D.R., Williams,L.C., de Malmarche,S.A. and
Bergquist,P.L.
A single emm gene-specific oligonucleotide probe does not recognise
all members of the Streptococcus pyogenes M type 1
JOURNAL FEMS Microbiol. Lett. 130 (2-3), 145-149 (1995)
MEDLINE 95377598
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq l69934] from the original journal article.
This sequence comes from Fig. 1.
Location/Qualifiers
1..229
/organism="Streptococcus pyogenes"
/db_xref="taxon:1314"
2..229
/partial
/gene="emm"
/notes="M protein"
2..229
/partial
/gene="emm"
/notes="This sequence comes from Fig. 2"
/codon_start=1
/product="M protein"
/protein_id="AAB35162.1"
/db_xref="GI:1087105"
/translation="SVAVRLTVLGFAGFANOTEVKANGDGNPREVIEDLAANPAIONI
RLRHNKDLKARLENAMVAGRDFKRAELEK"
BASE COUNT 84 a 32 c 59 g 54 t
ORIGIN
Query Match 13.0%; Score 151; DB 3; Length 229;
Best Local Similarity 100.0%; Pred. No. 2.4e-18;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 598 aacggtgtagtgaatccttaggaagttatagaagatcttgcagcaacaatcccgcaata 657
DB 65 AACGGTGTGTTAATCTCTAGGAAAGTTATAGAAGATCTTGACGCAACAATCCCGCAATA 124
QY 658 caaaatatactttacgtcagcaaaaacagagacttaaaagcgaattagagaatgcaatg 717
DB 125 CAAATATACGTTTACGTCACGAAACAAAGGACTTAAAGCGAGATTAGAGATGCAATG 184
QY 718 gaagttgaggaagaagattttaagagagctg 748
DB 185 GAAGTTGCGAGGAGAGATTTTAAGAGAGCTG 215
RESULT 10
S79254
LOCUS S79254 237 bp DNA BCT 16-MAY-1996
DEFINITION Streptococcus pyogenes M type 1 (emm1) gene, partial cds.
ACCESSION U11940
VERSION U11940.1 GI:533557
KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 237)
Whitmore,A.M.
Sequence Analysis of the Emm-Like Gene Family of Streptococcus
pyogenes
JOURNAL Thesis (1993) Microbiology, University of Newcastle Upon Tyne
REFERENCE 2 (bases 1 to 237)
Whitmore,A.M., Kapur,V., Sullivan,D.J., Musser,J.M. and Kehoe,M.A.
Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci
Mol. Microbiol. 14 (4), 619-631 (1994)
JOURNAL 95198537
MEDLINE
REFERENCE 3 (bases 1 to 237)
Whitmore,A.M.
Direct Submission
Submitted (07-JUL-1994) EMBL Data Library By: A. M. Whitmore,
Microbiology, University of Newcastle Upon Tyne, Framlington Place,
Newcastle Upon Tyne, UK, NE2 4HH
Location/Qualifiers
1..237
/organism="Streptococcus pyogenes"
/strain="M type 1"
```

gene	/isolate="NCTC8198"		/db_xref="taxon:1314"	
CDS	1..237		/gene="emml"	
	<1..>237		/gene="emml"	
	/codon_start=1		/transl_table=1	
	/protein_id="AAA9556.1"		/db_xref="GI:53358"	
	/translation="KTGPASVAVALTVLGAGFANOTEVKANGDNPREVIEDLAANP		/translation="KTGPASVAVALTVLGAGFANOTEVKANGDNPREVIEDLAANP	
	AQNTRLRHNKDKLARLENAMVAGRDFRAEEL"		AQNTRLRHNKDKLARLENAMVAGRDFRAEEL"	
BASE COUNT	86 a	35 c	62 g	54 t
ORIGIN				
Query Match	13.0%	Score 151;	DB 3;	Length 237;
Best Local Similarity	100.0%	Pred. No. 2.4e-18;		
Matches 151;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	598	aacggtgatgtaacctaggaagtattatagaagattcttcgacaaacaatccgcgaata	657	
DBD	79	AACGGTGATGCTAATCTCTAGGGAAGTTATAGAAGATCTTCACGACAAACAATCCGCAATA	138	
QY	658	caaaatatacgtttacgtccagcaaaacaaggacttaaacgagattagagaatgcaatg	717	
DBD	139	CAAAATATACGTTTACGTCACGACAAACAAGGACTTAAAGCGAGATTAGAGAATGCAATG	198	
QY	718	gaagttgcaggaagagattttaagagagctg	748	
DB	199	GAAGTTGCAGGAAGAGATTTTAAGAGAGCTG	229	
RESULT 11				
SP020094	SPU20094 300 bp DNA BCT 05-JAN-1996			
LOCUS	Streptococcus pyogenes M1.0 protein gene, partial cds.			
DEFINITION	U20094			
ACCESSION	U20094.1 GI:643545			
VERSION				
KEYWORDS	Streptococcus pyogenes.			
SOURCE	Streptococcus pyogenes			
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.			
REFERENCE	1 (bases 1 to 300)			
AUTHORS	Musser, J.M., Kapur, V., Szeto, J., Pan, X., Swanson, D.S. and Martin, D.R.			
TITLE	Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease			
JOURNAL	Infect. Immun. 63 (3), 994-1003 (1995)			
MEDLINE	95172752			
REFERENCE	2 (bases 1 to 300)			
AUTHORS	Swanson, D.S.			
TITLE	Direct Submission			
JOURNAL	Submitted (19-JAN-1995) Douglas S. Swanson, Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA			
FEATURES	Location/Qualifiers			
source	1..300			
	/organism="Streptococcus pyogenes"			
	/isolate="MGAS 2116"			
	/db_xref="taxon:1314"			
	<1..>300			
	/codon_start=1			
	/transl_table=1			
	/product="M1.0 protein"			
	/protein_id="AAA85107.1"			
	/db_xref="GI:643546"			
	/translation="TVLGAGFANOTEVKANGDNPREVIEDLAANP			
	/translation="TVLGAGFANOTEVKANGDNPREVIEDLAANP			
	KOLKARLENAMVAGRDFRAEELKAKALEIDQKDLTKLKLQDDYDLAKEST"			
BASE COUNT	126 a	45 c	67 g	62 t
ORIGIN				

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL MEDLINE COMMENT FEATURES Source	2 (bases 1 to 1494) Harbaugh,M.P., Podbielski,A., Huegel,S. and Cleary,P.P. Nucleotide substitutions lead to antigenic variation in the V-terminal non-repeat domains ofgroup A streptococci M1 protein Mol. Microbiol. 7 (1993) In press 3 (bases 1 to 1494) Harbaugh,M.P., Podbielski,A., Hugl,S. and Cleary,P.P. Nucleotide substitutions and small-scale insertion produce size and antigenic variation in group A streptococcal M1 protein Mol. Microbiol. 8 (5), 981-991 (1993) 93360826 Product subunit structure is a dimer. See also x62132 for sequence. Location/Qualifiers 1. .1494 /organism="Streptococcus pyogenes" /isolate="Prague 11-29/58" /db_xref="taxon:1314" 10. .14 /evidence=experimental 31. .1485 /gene="emm1" 31. .1485 /gene="emm1" /codon_start=1 /transl_table=11 /evidence=experimental /product="M protein type 1" /protein_id="CAA44062.1" /db_xref="GI:311758" /db_xref="SPTREMBL:Q10372" /translation="MAKNINHNHSLSKLLKTGTASVAVALTVLGAGFANGTEVVKANGD GNPREVDIAANNPAIDYQAIQRHNEDKLKARLENAMVAGDRFRAEELEEKAKQALE DQRDKDLTKLEQQDYDIASQSTSWQKLEKEEKKEALEIAIDQASGRDVRHRAE LESEKLEKKALEADQASQDYNNANVLEKLDITTRQENINRLNGNKKLEQLQS SEKQLTKTERAKLEEQKISASQSRSLRDLDSRAKKQVEKDLANLTAEIDKVKED KOISDASQGRSLRDLDSAREAKQVEKDLANLTAEIDKVKEEKQISDASQGRDLDD ASRAKKQVEKALEEANSKAALEKLNKELEESKLTEKEAKLEAQKALEANVLKDP LAKAQEELAKLRAGASQSDPTDKPKNKAVPGKGQPAQGTKPNQNKAPMKETKRL PSTGEATNPFTFAALTLYNATAGVAAVKRRKEEN"	624 a. 268 c 319 g 283 t
BASE COUNT		
ORIGIN		

Query Match	13.08;	Score 151;	DB 3;	Length 1494;	
Best local similarity	100.0%;	Pred. NO. 2.2e-18;			
Matches 151;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	598	aacggtgatgtaactctcagggaagttatagaagatcttcgacgaacaacatccgcgaata	657		
Db	154	RACSGTGATGGTAATCCTTAGGGAAGTTATAGAAGATCTTGACGACAAACAAATCCGCAATA	213		
QY	658	caaaatatacgtttacgtgcacgaacaacaggacctaaagcggagattagagaagaatcgatg	717		
Db	214	CAAAATATACGTTTAGTGCACGNAACANGAGACTTAAAGCGAGATTAGAGAATGCAATG	273		
QY	718	gaagttgcaggaagagattttaagagagctg	748		
Db	274	GAAGTTGCAGGAAGAGATTTTAAAGAGAGCTG	304		
RESULT	14				
SPENMI					
LOCUS	SPENMI	1708 bp	DNA	BC1	14-NAR-1996
DEFINITION	S.pyogenes emml gene for type I group A M protein.				
ACCESSION	X07933 X07860				
VERSION	X07933.1 GI:47358				
KEYWORDS	emml gene; M protein.				
SOURCE	Streptococcus pyogenes.				
ORGANISM	Streptococcus pyogenes				
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
AUTHORS	1 (bases 1 to 1708) Hannes-Fritz,E.				

TITLE	Direct Submission
JOURNAL	Submitted (15-JUN-1988) Hannes-Fritz E., University of Minnesota, Department of Microbiology, Box 196, Mayo Building, Minneapolis, MN 55455
REFERENCE	2 (bases 1 to 1708)
AUTHORS	Hannes-Fritz,E., Kraus,W., Burdett,V., Dale,J.B., Beachey,E.H. and Cleary,P.
TITLE	Comparison of the leader sequences of four group A streptococcal M protein genes
JOURNAL	Nucleic Acids Res. 16 (10), 4667-4677 (1988)
MEDLINE	88247768
FEATURES	Location/Qualifiers
source	1..1708
	/organism="Streptococcus pyogenes"
	/strain="CS130"
	/db_xref="taxon:1314"
	/clone="pbf82"
	921..963
promoter	/note="pot. promoter region"
RS	1015..1019
CDS	1028..>1708
	/codon_start=1
	/transl_table=11
	/product="M protein precursor"
	/protein_id="CAA30757.1"
	/db_xref="GI:47359"
	/db_xref="SPTREMBL:Q10371"
	/translation="NAKNTNRRHSLKLTGTASVAVALTVLGAGFANOTEVKANGDGNPREVDLAANNAQINILRRHNKDKLARNENAVAGRDFKRAEELEKAKOALE DOKDELTKLKELOQDYDLAKESDSDROBLEKELEKEKALELAIDQASRDYHRATA LEKELEKEKALELAIDQASDYNNRVNLEKELEITITREQEINRNLLGNLKNLKLDELQLS SEKEQTIR
sig_peptide	1028..1150
mat_peptide	1151..>1708
repeat_region	/product="mature M protein, type 1 group A"
repeat_region	1427..1525
	/rpt_type=DIRECT
	1522..1609
	/rpt_type=DIRECT
BASE COUNT	640 a 310 c 331 g 427 t
ORIGIN	
Query Match	13.0%; Score 151; DB 3; Length 1708;
Best Local Similarity	100.0%; Pred. No. 2.2e-18;
Matches 151; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Qy	598 aacggtagtggaatcctcctagggaagttagaagatcttcgacgaacaacatccgcgaata 657
Ddb	1151 AACGGTGATGGTAACTCCTAGGGAAGTATAGAAGATCTTGAGCAAAACAAATCCGCAATA 1210
Qy	658 caaataatcgtttacgtcacgaagaacacaggagcttaaacgagagattagagaatgcaatg 717
Ddb	1211 CAAAATATACGTTTACGTACGCAAAACAGGACTTTAAAGCGGAGATTAGAGAATGCAATG 1270
Qy	718 gaagttgcaggaagagattttaagagagctg 748
Ddb	1271 GAAGTTGCAGGAAGAGATTTTAAAGAGAGCTG 1301

RESULT	15
SPU11997	
LOCUS	251 bp DNA BCT 16-MAY-1996
DEFINITION	Streptococcus pyogenes M type 68 (emml) gene, partial cds.
ACCESSION	U11997
VERSION	U11997.1 GI:533671
KEYWORDS	.
SOURCE	Streptococcus pyogenes.
ORGANISM	Streptococcus pyogenes
	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
REFERENCE	Streptococcus.
AUTHORS	1 (bases 1 to 251)
	Whatmore, A.M.

TITLE Sequence Analysis of the Emm-Like Gene Family of Streptococcus
Pyogenes
JOURNAL Thesis (1993) Microbiology, University of Newcastle Upon Tyne
REFERENCE 2 (bases 1 to 251)
AUTHORS Whatmore,A.M., Kapur,V., Sullivan,D.J., Musser,J.M. and Kehoe,M.A.
TITLE Non-congruent relationships between variation in emm gene sequences
and the population genetic structure of group A streptococci
JOURNAL Mol. Microbiol. 14 (4), 619-631 (1994)
MEDLINE 95198537
REFERENCE 3 (bases 1 to 251)
AUTHORS Whatmore,A. M.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1994) Embl Data Library By: A. M. Whatmore,
Microbiology, University of Newcastle Upon Tyne, Framlington Place,
Newcastle Upon Tyne, UK, NE2 4HH
FEATURES
source
1..251
Location/Qualifiers
/organism="Streptococcus pyogenes"
/strain="M type 68"
/isolate="R80/2076"
/db_xref="taxon:1314"
1..251
/gene="emm1"
<1..>251
/gene="emm1"
/codon_start=1
/transl_table=11
/protein_id="AA099613.1"
/db_xref="GI:533672"
/translation="LRKLKGTASVAVALTVLGAFANQTEYKANGDGNPREVIEDLP
ANPAIQIRLHNKDLKARLENAMEVAGRDFKRAEELS"
BASE COUNT 92 a 38 c 62 g 59 t
ORIGIN

Query Match 12.9%; Score 149.4; DB 3; Length 251;
Best Local Similarity 99.3%; Pred. No. 4.8e-18;
Matches 150; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 598 aacgggtgatcctcctaggggaagttatagaagatcttgcagcaacaatcccgcaata 657
|||||
DB 91 AACGGTGATGTAATCCTAGGGAAGTTATAGAAGATCTTCCAGCAACAATCCCGCAATA 150
|||||
QY 658 caaaatacacgtttacgtcacgaaacagagacttaaaagcgagattagagaatgcaatg 717
|||||
DB 151 CAAAATATACGTTTACGTCACGAAACAAAGGACTTAAAGCGAGATTAGAGAAATGCAATG 210
|||||
QY 718 gaagttgcagagagagattttaagaagctg 748
|||||
DB 211 GAAGTTGCAGGAGAGATTTTAAAGAGCTG 241
|||||

Search completed: May 13, 2001, 04:55:34
Job time: 10569 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 02:42:00 ; Search time 119.14 Seconds
(without alignments)
5674.154 Million cell updates/sec

Title: US-09-151-409-15

Perfect score: 1158

Sequence: 1 gcatcgatgtcgcgactag.....aggcagtaagtaaaagctt 1158

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_0401.*

1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1158	100.0	1158	20	X33103
2	419	36.2	741	21	A57894
3	417.4	36.0	741	15	Q45215
4	328.2	28.3	822	15	Q45216
5	328.2	28.3	822	21	A57895
6	301.8	26.1	741	15	Q45219
7	301.8	26.1	741	21	A57898
8	301.8	26.1	1029	15	Q45221
9	301.8	26.1	1029	21	A57900
10	241	20.8	861	15	Q45214
11	241	20.8	861	21	A57893

12	239.4	20.7	522	15	Q45220
13	239.4	20.7	522	21	A57899
14	124.4	10.7	561	15	Q45217
15	124.4	10.7	561	21	A57896
16	106.4	9.2	1617	18	T49317
17	105.6	9.1	301	6	N50341
18	76.2	6.6	408	15	Q45223
19	76.2	6.6	408	21	A57902
20	68	5.9	2943	21	A70229
21	64.2	5.5	3579	21	A70099
22	63.2	5.5	1998	21	A70212
23	61	5.3	6644	20	X33181
24	61	5.3	7372	20	X33182
25	61	5.3	7797	20	X33180
26	61	5.3	7996	20	X33184
27	57.6	5.0	1686	16	Q87587
28	56.4	4.9	1039	20	X61812
29	56.4	4.9	1125	20	X61811
30	56.4	4.9	35515	20	X20252
31	55.8	4.8	3399	17	T05868
32	54.8	4.7	855	15	Q45161
33	53.6	4.6	7458	21	A70106
34	53.2	4.6	204	15	Q45222
35	53.2	4.6	204	21	A57901
36	53.2	4.6	918	15	Q45218
37	53.2	4.6	918	21	A57897
38	53.2	4.6	2658	21	A70213
39	53	4.6	765	15	Q45160
40	52.4	4.5	3738	21	A70178
41	51.8	4.5	4590	7	M60472
42	50.6	4.4	7503	21	A70206
43	49.2	4.2	5940	21	A70105
44	49	4.2	1664976	19	V21209
45	48.4	4.2	11922	21	A70187

ALIGNMENTS

RESULT 1
X33103
ID X33103 standard; DNA; 1158 BP.
XX
AC X33103;
XX
DT 23-JUN-1999 (first entry)
XX
DE Streptococcal hexavalent M protein vaccine encoding DNA.
XX
KW Streptococcal; hexavalent emm gene; multivalent; vaccine; Group A;
KW immunogenic; immune response; pathogen; fusion protein; ss.
XX
OS Streptococcus sp.
XX
PN WO9913084-A1.
XX
PD 18-MAR-1999.
XX
PF 14-SEP-1998; 98WO-US19100.
XX
PR 12-SEP-1997; 97US-0058635.
XX
(IDVA-) ID VACCINE.
XX
Dale JB;
XX
WPI; 1999-215066/18.
XX
P-PSDB; Y04368.
XX
Immunogenic fusion protein derived from group A streptococci
XX
Example 1; Fig 7; 50pp; English.

Recombinant M24-M5
S. pyogenes hybrid
Recombinant M24-M5
S. pyogenes hybrid
Type-6 M-protein e
Sequence of a port
M24-M5-M6-M19-M3-M
S. pyogenes octava
Plasmodium falcipa
Plasmodium falcipa
Plasmodium falcipa
Base sequence of t
Base sequence of t
Cowpox virus bsr f
Base sequence of t
DNA encoding leuco
B. burgdorferi ant
B. burgdorferi ant
Borrelia burgdorfe
Chicken leucocytoz
Sequence of fragme
Plasmodium falcipa
M24-M5-M6-M19 hybr
S. pyogenes hybrid
Recombinant M24-M5
S. pyogenes hybrid
Plasmodium falcipa
Sequence of a frag
Plasmodium falcipa
Sequence encoding
Plasmodium falcipa
Plasmodium falcipa
Methanococcus jann
Plasmodium falcipa

CC The present invention describes an immunogenic fusion polypeptide (I)
CC that stimulates an immune response against group A streptococci (GAS).
CC (I) comprises: (a) at least 2 peptides (Ia) from GAS, at least 10 amino
CC acids long and able to induce a specific immune response; and (b) a
CC peptide (Ib), C-terminal to (Ia) that protects the immunogenicity of the
CC component described in (a) but is not essential for stimulation of the
CC immune response. Vaccines containing (I) are used to protect against GAS
CC infections, specifically those caused by Streptococcus pyogenes, e.g.
CC pharyngitis, pyoderma, toxic shock syndrome, deep tissue invasion, sepsis
CC and acute rheumatic fever. (I) have improved immunogenicity and do not
CC generate antibodies that are cross-reactive with human tissues. The
CC present sequence encodes a hexavalent M protein vaccine.
XX
SQ Sequence 1158 BP; 474 A; 165 C; 255 G; 264 T; 0 other;

Query Match 100.0%; Score 1158; DB 20; Length 1158;
Best Local Similarity 100.0%; Pred. No. 4,1e-244;
Matches 1158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatgcatgttcgactaggtctcagacagatactctggaagagtagcaagaacgtgct 60
Db 1 gcatgcatgttcgactaggtctcagacagatactctggaagagtagcaagaacgtgct 60
QY 61 gacaagttgagatagaaaaacaatacgtttaaaacttaagaatagtagtgaattttaat 120
Db 61 gacaagttgagatagaaaaacaatacgtttaaaacttaagaatagtagtgaattttaat 120
QY 121 aataaagcgttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagag 180
Db 121 aataaagcgttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagag 180
QY 181 aaactacgttgatccgcgtgactaggggtacataataatgacccgcaagaagcaaaagaa 240
Db 181 aaactacgttgatccgcgtgactaggggtacataataatgacccgcaagaagcaaaagaa 240
QY 241 gctcttacaagtagtagctagaaaaccatgacttaaaactaaagaatgaaggggtaaaaa 300
Db 241 gctcttacaagtagtagctagaaaaccatgacttaaaactaaagaatgaaggggtaaaaa 300
QY 301 actgagaatgaaggggttaaaactggaatgaaggggttaaaactgagaatgaaggggtta 360
Db 301 actgagaatgaaggggttaaaactggaatgaaggggttaaaactgagaatgaaggggtta 360
QY 361 aaactgaggtcagacagaggtttctcaggggacgtgagaaacccgcagacaaacagca 420
Db 361 aaactgaggtcagacagaggtttctcaggggacgtgagaaacccgcagacaaacagca 420
QY 421 gaactcttaacaagtagtagcgttagagaactctatgtttacaagctaaataatgacaagtta 480
Db 421 gaactcttaacaagtagtagcgttagagaactctatgtttacaagctaaataatgacaagtta 480
QY 481 ccattgagagtgctttactactagcagcagcagcagcagcagcagcagcagcagcagcagc 540
Db 481 ccattgagagtgctttactactagcagcagcagcagcagcagcagcagcagcagcagcagc 540
QY 541 gatcttgacgcaaaagaacatgaattacaacacagacagacagacagacagacagacagac 600
Db 541 gatcttgacgcaaaagaacatgaattacaacacagacagacagacagacagacagacagac 600
QY 601 ggtgatggttaactcctagggaggttatagaagatttttcaagaacacacacacacacacac 660
Db 601 ggtgatggttaactcctagggaggttatagaagatttttcaagaacacacacacacacacac 660
QY 661 aataacgtttacgtcacgcaaaacagacgttaaaagcagagattagagaaatgcaatggaa 720
Db 661 aataacgtttacgtcacgcaaaacagacgttaaaagcagagattagagaaatgcaatggaa 720
QY 721 gtgcaggagaagattttaagagagctggttaccctgttagatcagaggttacacattatat 780
Db 721 gtgcaggagaagattttaagagagctggttaccctgttagatcagaggttacacattatat 780
QY 781 actaaacataatagtaattaccacaacataataatgacaaagctggcagacttgacctgaga 840

Db 781 actaaacataatagtaattaccacaacataataatgacaaagctggcagacttgacctgaga 840
QY 841 caaagcgtgaatatactaaagcgttaataatgattggcgtgagagcgtgttacaagagttta 900
Db 841 caaagcgtgaatatactaaagcgttaataatgattggcgtgagagcgtgttacaagagttta 900
QY 901 aatcagatgctcgactaggtctcagacagatactctggaagagtagcaagaacgtgct 960
Db 901 aatcagatgctcgactaggtctcagacagatactctggaagagtagcaagaacgtgct 960
QY 961 gacaagttgagatagaaaaacaatacgtttaaaacttaagaatagtagtgaattttaat 1020
Db 961 gacaagttgagatagaaaaacaatacgtttaaaacttaagaatagtagtgaattttaat 1020
QY 1021 aataaagcgttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagag 1080
Db 1021 aataaagcgttaaaagatcataatgatgagtttaactgaagagttgagtaactgaagag 1080
QY 1081 aaactacgttaaaatgataaatcactatctgaaaaagctagtaaaattcaagaattagag 1140
Db 1081 aaactacgttaaaatgataaatcactatctgaaaaagctagtaaaattcaagaattagag 1140
QY 1141 gcacgttaagtaaaagcctt 1158
Db 1141 gcacgttaagtaaaagcctt 1158
RESULT 2
A57894
ID A57894 standard; DNA; 741 BP.
XX A57894;
XX
XX
DT 10-OCT-2000 (first entry)
XX
DE S. pyogenes hybrid M protein (M24-M5-M6-M19) DNA, SEQ ID NO:3.
XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia; ds.
XX Streptococcus pyogenes.
OS Synthetic.
XX
XX US6063386-A.
XX
XX 16-MAY-2000.
XX
XX 15-SEP-1997; 97US-0937271.
XX
XX 16-SEP-1992; 92US-0945954.
XX
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX
XX Lederer JW, Dale JB;
XX
XX WPI; 2000-364475/31.
DR P-PSDB; B031114.
XX
XX New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever -
XX Disclosure; Fig 4A-B; 62pp; English.
XX
XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein

Sequence 741 BP; 311 A; 106 C; 164 G; 160 T; 0 other;

XX	antibody; heart tissue; antigen; serotype; mucosal; ss.
XX	
OS	Synthetic.

Query Match	36.0%;	Score 417.4;	DB 15;	Length 741;
Best Local Similarity	78.8%;	Pred. No. 1.7e-82;		
Matches 584:	Conservative	0:	Mismatches 1:	Indels 156:

1.

Tue May 15 07:27:08 2001

PN	WO9406421-A.	XX	31-MAR-1994.	XX	31-MAR-1994.
PD		XX		XX	
PF	15-SEP-1993;	XX	93WO-US08703.	XX	15-SEP-1993;
PR	16-SEP-1992;	XX	92US-0945954.	XX	16-SEP-1992;
PA	(UYTE-) UNIV TENNESSEE RES CORP.	XX		XX	
PI	Dale JB, Lederer JW;	XX		XX	
DR	WPI; 1994-118122/14.	XX		XX	
DR	P-PSDB; R50994.	XX		XX	
XX	New immunogenic hybrid proteins derives from streptococcal M proteins	XX		XX	
PT	- induces opsonic antibodies, for protective immunisation against	XX		XX	
PT	against multiple group A streptococci serotypes	XX		XX	
PS	Disclosure; Fig 6; 67pp; English.	XX		XX	
XX	The sequences given in Q45214-23 encode hybrid M proteins which	XX		XX	
CC	contain the M24-M5-M6 and/or M19 subunits. These tetraivalent	CC		CC	
CC	proteins were constructed using fragments of the 5' regions of emm	CC		CC	
CC	genes that were amplified by PCR, ligated in tandem and expressed in	CC		CC	
CC	pkK223.3. The amplified regions pref. encode protective and not	CC		CC	
CC	tissue-cross-reactive epitopes, which can then be linked into one	CC		CC	
CC	protein molecule. The recombinant hybrid protein may contain 113	CC		CC	
CC	N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and	CC		CC	
CC	35 from M19. Each section is linked by 2 amino acids specified by	CC		CC	
CC	the respective restriction enzyme sites that were synthesised into	CC		CC	
CC	the primers used to specify the PCR product. Multivalent M proteins	CC		CC	
CC	such as this may be used for protective immunisation against group A	CC		CC	
CC	streptococci, which esp. cause rheumatic fever and rheumatic heart	CC		CC	
CC	disease. Humoral antibodies raised against these proteins do not	CC		CC	
CC	react with heart tissue antigens but are effective against many	CC		CC	
CC	different serotypes. The multivalent proteins may also include	CC		CC	
CC	sequences which induce mucosal antibodies and do not require coupling	CC		CC	
CC	to an immunogenic carrier.	CC		CC	
SQ	Sequence 822 BP; 332 A; 136 C; 182 G; 172 T; 0 other;	SQ		SQ	
Query Match	28.3%; Score 328.2; DB 15; Length 822;	Query Match	28.3%; Score 328.2; DB 15; Length 822;	Query Match	28.3%; Score 328.2; DB 15; Length 822;
Best Local Similarity	87.6%; Pred. No. 5.1e-63;	Best Local Similarity	87.6%; Pred. No. 5.1e-63;	Best Local Similarity	87.6%; Pred. No. 5.1e-63;
Matches	404; Conservative 0; Mismatches 3; Indels 54; Gaps	Matches	404; Conservative 0; Mismatches 3; Indels 54; Gaps	Matches	404; Conservative 0; Mismatches 3; Indels 54; Gaps
Qy	185 tacgtgacccgctgactagggtacaataaataatgacccgcaaaagagcaaaagagctc	Qy	185 tacgtgacccgctgactagggtacaataaataatgacccgcaaaagagcaaaagagctc	Qy	185 tacgtgacccgctgactagggtacaataaataatgacccgcaaaagagcaaaagagctc
Db	362 ttccaggatcccgctgactagggtacaataaataatgacccgcaaaagagcaaaagagctc	Db	362 ttccaggatcccgctgactagggtacaataaataatgacccgcaaaagagcaaaagagctc	Db	362 ttccaggatcccgctgactagggtacaataaataatgacccgcaaaagagcaaaagagctc
Qy	245 ttgacaagtatgactagaaaaccatgacttaaaataagaatgaagggttaaaaaactg	Qy	245 ttgacaagtatgactagaaaaccatgacttaaaataagaatgaagggttaaaaaactg	Qy	245 ttgacaagtatgactagaaaaccatgacttaaaataagaatgaagggttaaaaaactg
Db	422 ttgacaagtatgactagaaaaccatgacttaaaataagaatgaagggttaaaaaactg	Db	422 ttgacaagtatgactagaaaaccatgacttaaaataagaatgaagggttaaaaaactg	Db	422 ttgacaagtatgactagaaaaccatgacttaaaataagaatgaagggttaaaaaactg
Qy	305 agaataaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaa	Qy	305 agaataaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaa	Qy	305 agaataaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaa
Db	482 agaataaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaa	Db	482 agaataaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaa	Db	482 agaataaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaa
Qy	365 ctgag-----gtcgacagagtggttctctaggggggacgg	Qy	365 ctgag-----gtcgacagagtggttctctaggggggacgg	Qy	365 ctgag-----gtcgacagagtggttctctaggggggacgg
Db	542 ctgaggtcgcaccaggaaacccagctgttcacgctcagacagagtggttctctaggggggacgg	Db	542 ctgaggtcgcaccaggaaacccagctgttcacgctcagacagagtggttctctaggggggacgg	Db	542 ctgaggtcgcaccaggaaacccagctgttcacgctcagacagagtggttctctaggggggacgg
Qy	398 tagaaaaccgggacaaagcagcagaaacttttaacaagtatgacgtagagaactatgt	Qy	398 tagaaaaccgggacaaagcagcagaaacttttaacaagtatgacgtagagaactatgt	Qy	398 tagaaaaccgggacaaagcagcagaaacttttaacaagtatgacgtagagaactatgt
Db	602 tagaaaaccgggacaaagcagcagaaacttttaacaagtatgacgtagagaactatgt	Db	602 tagaaaaccgggacaaagcagcagaaacttttaacaagtatgacgtagagaactatgt	Db	602 tagaaaaccgggacaaagcagcagaaacttttaacaagtatgacgtagagaactatgt
Qy	458 tacaagctaataatgacaagttacc-----atggagag	Qy	458 tacaagctaataatgacaagttacc-----atggagag	Qy	458 tacaagctaataatgacaagttacc-----atggagag
Db	662 tacaagctaataatgacaagttacc-----atggagag	Db	662 tacaagctaataatgacaagttacc-----atggagag	Db	662 tacaagctaataatgacaagttacc-----atggagag
Qy	491 tgcgttactagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	Qy	491 tgcgttactagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	Qy	491 tgcgttactagggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc

Db 722 tgcgttatactaggccataccgagaataagctaaataattattgacgatcttgacg 781
QY 551 caaaagaacatgaattacacacagaatgaagaattatct 591
Db 782 caaaagaacatgaattacacacagaatgaagaattatct 822

RESULT 5
A57895
ID A57895 standard; DNA; 822 BP.
AC A57895;
XX
XX
10-OCT-2000 (first entry)
DE
DE S. pyogenes hybrid M protein (M24-M5-M6-M19) DNA, SPQ ID NO:5.
XX
KW Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
XX pneumonia; ds.
XX
OS Streptococcus pyogenes.
OS Synthetic.
XX
XX US06063386-A.
XX
XX 16-MAY-2000.
XX
XX 15-SEP-1997; 97US-0937271.
XX
XX 16-SEP-1992; 92US-0945954.
XX
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX
XX Lederer JW, Dale JB;
XX
XX WPI; 2000-364475/31.
XX
XX P-PSDB; B03115.
XX
XX New immunogenic recombinant hybrid M protein comprising amino-terminal
XX peptide fragments of streptococcal M protein useful as vaccine against
XX rheumatic fever and infections leading to rheumatic fever -
XX
XX Disclosure; Fig 6A-B; 62pp; English.
XX
XX The invention relates to multivalent immunogenic hybrid group A
XX streptococcal M proteins comprising N-terminal peptide fragments of M
XX proteins that elicit opsonic antibodies against multiple serotypes
XX of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
XX generated using the hybrid proteins are against one or more M protein
XX serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
XX The invention also encompasses a recombinant DNA molecule comprising a
XX nucleotide sequence that encodes a multivalent hybrid M protein; and a
XX method for immunising a mammal against streptococcal infections,
XX comprising administering an immunogenic multivalent hybrid M protein to
XX the mammal. The multivalent hybrid M proteins are useful for eliciting
XX opsonic or protective antibodies to the M proteins of different serotypes
XX of group A streptococci, and may therefore be used as vaccines to protect
XX against and control infection by type A streptococci. Type A streptococci
XX are not only responsible for streptococcal pharyngitis (strep throat),
XX forms of pneumonia and a condition resembling toxic shock, but are also
XX involved in the development of acute rheumatic fever (ARF) and rheumatic
XX heart disease. In a patient with ARF, antibodies formed during a group A
XX streptococcal infection are also cross-reactive with heart tissue, which
XX indicates that the streptococci and host tissue contain similar antigenic
XX motifs. The new multivalent vaccines are capable of raising sero-specific
XX antibodies against various serotypes of group A streptococci which are
XX not cross-reactive with human heart tissue. Sequences A57893-A57902
XX represent DNAs encoding multivalent hybrid Streptococcus pyogenes M
XX proteins generated in the disclosure of the invention.

SQ Sequence 822 BP; 332 A; 136 C; 182 G; 172 T; 0 other;
Query Match 28.3%; Score 328.2; DB 21; Length 822;
Best Local Similarity 87.6%; Pred. No. 5.1e-63;
Matches 404; Conservative 0; Mismatches 3; Indels 54; Gaps 2;
QY 185 tacgtggtatccgctgactaggggtacataataatgaccgcgaagaagcgaagcctc 244
Db 362 ttccaggatccgctgactaggggtacataataatgaccgcgaagaagcgaagcctc 241
QY 245 ttgacaagtatgactagaaacccatgacttaaaactagaatgaagggttaaaaactg 304
Db 422 ttgacaagtatgactagaaacccatgacttaaaactagaatgaagggttaaaaactg 481
QY 305 agaataagggttaaaaactgagaatgaagggttaaaaactgagaatgaagggttaaaaactg 364
Db 482 agaataagggttaaaaactgagaatgaagggttaaaaactgagaatgaagggttaaaaactg 541
QY 365 ctgag-----gtcacagagtggttcttaggggacgg 397
Db 542 ctgaggtcgacccagggaacccagctgtccagtcgacagagtggttcttaggggacgg 601
QY 398 tagaaaacccggaacaaagcagagacttttaacaagtatgacgtagagaactctatgt 457
Db 602 tagaaaacccggaacaaagcagagacttttaacaagtatgacgtagagaactctatgt 661
QY 458 tacaagctaaatgacaagttacc-----atggagag 490
Db 662 tacaagctaaatgacaagttaccatggccagggaacccagctgttccacatgagag 721
QY 491 tgcgttatactaggccataccgagaataagctaaataattattgacgatcttgacg 550
Db 722 tgcgttatactaggccataccgagaataagctaaataattattgacgatcttgacg 551
QY 551 caaaagaacatgaattacacacagaatgaagaattatct 591
Db 782 caaaagaacatgaattacacacagaatgaagaattatct 822

RESULT 6
Q45219
ID Q45219 standard; DNA; 741 BP.
XX
XX AC Q45219;
XX
XX DT 02-NOV-1994 (first entry)
XX
XX DE Recombinant M19-M6-M5-M24 tetraivalent hybrid gene.
XX
XX KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
XX M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
XX tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
XX restriction enzyme site; multivalent M protein; immunisation; group A;
XX streptococci; rheumatic fever; rheumatic heart disease; humoral;
XX antibody; heart tissue; antigen; serotype; mucosal; ss.
XX Synthetic.
XX
XX OS WO9406421-A.
XX
XX PN 31-MAR-1994.
XX
XX PD 15-SEP-1993; 93WO-US08703.
XX
XX PF 16-SEP-1992; 92US-0945954.
XX
XX PR (UYTE-) UNIV TENNESSEE RES CORP.
XX
XX PA Dale JB, Lederer JW;
XX
XX PI WPI; 1994-118122/14.
XX
XX DR P-PSDB; R50997.
XX

CC proteins generated in the disclosure of the invention.
 XX Sequence 741 BP; 311 A; 107 C; 163 G; 160 T; 0 other;
 SQ

Query Match 26.1%; Score 301.8; DB 21; Length 741;
 Best Local Similarity 68.3%; Pred. No. 2.9e-57;
 Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;

QY 486 gagagtcgttatactaggcaccagagataagctaaagcttaagagattgagagctgtc 545
 |||||
 Db 3 gagagtcgttatactaggcaccagagataagctaaagcttaagagattgagagctgtc 62
 |||||
 QY 546 tgacgcaaaagacatgaattacaacacagaaatgagaagttatctctgcagacggtga 605
 |||||
 Db 63 tgacgcaaaagacatgaattacaacacagaaatgagaagttatctctgcagacggtgtt 122
 |||||
 QY 606 tggtaatccttagggaagttatagaaatcttgacgaaacaatcccgcaatacaaaatat 665
 |||||
 Db 123 tcttaggggacggtagaaaaccgg-----acaaagc 155
 |||||
 QY 666 acgtttacgtcaacgaaacagagacttaaaagcgagattgagagaaatgcaatggaagtgc 725
 |||||
 Db 156 acgagaactctttaacaagtatgacgtaggaactctatgttacaagcttaataagacaa 215
 |||||
 QY 726 aggaagagattttaacacaaataatgacaaagctggcagacttgacattacacaattatataa 785
 |||||
 Db 216 cttagtcgacgcgtgactaggggtacaataatgacccgcaaaagcaaaagagctct 275
 |||||
 QY 786 acataatagtaattaccaacaataatgacaaagctggcagacttgacattacacaattatataa 845
 |||||
 Db 276 tgacaaagtatgagctagaaacacatgacttaaaactaagaatgaagggttaaaactga 335
 |||||
 QY 846 ggctgaatatctaaagcccttaattggtggtgagagctgttacaagagtttaaatat 905
 |||||
 Db 336 gaatgaagggttaaaactgagaatgaagggtttaaactgagaatgaagggttaaaact 395
 |||||
 QY 906 cga-----tgccgactaggctcagacagatactctggaagaaagctacaagacgtgc 959
 |||||
 Db 396 tgagccatggttcgactaggctcagacagatactctggaagaaagctacaagacgtgc 455
 |||||
 QY 960 tgacaagttgagatagaaaacactgttaaaacttaagaatagctacttaagtttaa 1019
 |||||
 Db 456 tgacaagttgagatagaaaacactgttaaaacttaagaatagctacttaagtttaa 515
 |||||
 QY 1020 taataaagcgttaaaagatcataatgatgagtttaactgaaaggttgagtaagtctaaga 1079
 |||||
 Db 516 taataaagcgttaaaagatcataatgatgagtttaactgaaaggttgagtaagtctaaga 575
 |||||
 QY 1080 gaaactacgtataaataatgataatcactatctgaaaaagctgtaaaattcaagaattaga 1139
 |||||
 Db 576 gaaactacgtataaataatgataatcactatctgaaaaagctgtaaaattcaagaattaga 635
 |||||
 QY 1140 ggcacgtaaagtaaaagctt 1158
 |||||
 Db 636 ggcacgtaaagctgactctt 654

RESULT 8

Q45221
 ID Q45221 standard; DNM; 1029 BP.
 XX
 AC Q45221;
 XX
 DT 02-NOV-1994 (first entry)
 DE Tetraivalent-C repeat gene.
 XX
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; pKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; humoral;

KW antibody; heart tissue; antigen; serotype; mucosal; ss.
 XX Synthetic.
 OS
 PN WO9406421-A.
 XX
 PD 31-MAR-1994.
 XX
 PF 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Dale JB, Lederer JW;
 XX
 DR WPI; 1994-118122/14.
 XX
 P-PSDB; R50999.
 XX
 PS New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX
 PS Disclosure; Fig 11; 67pp; English.

The sequences given in Q45214-23 encode hybrid M proteins which contain the M24-M5-M6 and/or M19 subunits. These tetraivalent proteins were constructed using fragments of the 5' regions of emm genes that were amplified by PCR, ligated in tandem and expressed in pKK223.3. The amplified regions pref. encode protective and not tissue-cross-reactive epitopes, which can then be linked into one protein molecule. The recombinant hybrid protein may contain 113 N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and 35 from M19. Each section is linked by 2 amino acids specified by the respective restriction enzyme sites that were synthesised into the primers used to specify the PCR product. Multivalent M proteins such as this may be used for protective immunisation against group A streptococci, which esp. cause rheumatic fever and rheumatic heart disease. Humoral antibodies raised against these proteins do not react with heart tissue antigens but are effective against many different serotypes. The multivalent proteins may also include sequences which induce mucosal antibodies and do not require coupling to an immunogenic carrier.

SQ Sequence 1029 BP; 424 A; 168 C; 226 G; 211 T; 0 other;

Query Match 26.1%; Score 301.8; DB 15; Length 1029;
 Best Local Similarity 68.3%; Pred. No. 3.1e-57;
 Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;

QY 486 gagagtcgttatactaggcaccagagataagctaaagcttaagagattgagagctgtc 545
 |||||
 Db 3 gagagtcgttatactaggcaccagagataagctaaagcttaagagattgagagctgtc 62
 |||||
 QY 546 tgacgcaaaagacatgaattacaacacagaaatgagaagttatctctgcagacggtga 605
 |||||
 Db 63 tgacgcaaaagacatgaattacaacacagaaatgagaagttatctctgcagacggtgtt 122
 |||||
 QY 606 tggtaatccttagggaagttatagaaatcttgacgaaacaatcccgcaatacaaaatat 665
 |||||
 Db 123 tcttaggggacggtagaaaaccgg-----acaaagc 155
 |||||
 QY 666 acgtttacgtcaacgaaacagagacttaaaagcgagattgagagaaatgcaatggaagtgc 725
 |||||
 Db 156 acgagaactctttaacaagtatgacgtaggaactctatgttacaagcttaataagacaa 215
 |||||
 QY 726 aggaagagattttaacagagagctgactgtttagatcaggtttacacaattatataa 785
 |||||
 Db 216 cttagtcgacgcgtgactaggggtacaataatgacccgcaaaagcaaaagagctct 275
 |||||
 QY 786 acataatagtaattaccaacaataatgacaaagctggcagacttgagacactgaacaaa 845
 |||||

Db 276 tgacaagtatgaagtagaagaaacacatgacttaataaaactaagaatgaagggttaaaaaactga 335
 QY 846 ggctgaatatctaaagcccttaattgattggctgagaggctgttacaagagttaaatat 905
 Db 336 gaatgaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttaaaaaac 395
 QY 906 cga-----tgfcgcgactagctcagacagatactctggaagaaagtaacaagaactgc 959
 Db 396 tgagccatgggttcgcgactagctcagacagatactctggaagaaagtaacaagaactgc 455
 QY 960 tgacaagtttgagatagaagaacaatcagtttaaaacttaagaatgaagtagacttaatttaa 1019
 Db 456 tgacaagtttgagatagaagaacaatcagtttaaaacttaagaatgaagtagacttaatttaa 515
 QY 1020 taataaagcgttaaaagatcataatgatgagttaaactgaagtgtagtaatgcttaaga 1079
 Db 516 taataaagcgttaaaagatcataatgatgagttaaactgaagtgtagtaatgcttaaga 575
 QY 1080 gaaactacgtataaaatgataaaatcactatctgaaaaagctagtaaaatcaagaattaga 1139
 Db 576 gaaactacgtataaaatgataaaatcactatctgaaaaagctagtaaaatcaagaattaga 635

QY 1140 ggcacgttaagtaaaagctt 1158

Db 636 ggcacgttaagtgatctt 654

RESULT 9

A57900

ID A57900 standard; DNA; 1029 BP.

XX AC A57900;

XX DT 10-OCT-2000 (first entry)

XX DE S. pyogenes hybrid M protein (M19-M6-M5-M24) DNA, SEQ ID NO:16.

XX KW Multivalent hybrid M protein; group A streptococcus; serotype; immunogenic; sero-specific antibody; streptococcal infection;

KW cross reactivity; vaccine; acute rheumatic fever; ARF;

KW rheumatic heart disease; streptococcal pharyngitis; strep throat;

KW pneumonia; ds.

XX OS Streptococcus pyogenes.

OS Synthetic.

XX PN US6063386-A.

XX PD 16-MAY-2000.

XX PF 15-SEP-1997; 97US-0937271.

XX PR 16-SEP-1992; 92US-0945954.

XX PA (UYTE-) UNIV TENNESSEE RES CORP.

XX PI Lederer JW, Dale JB;

XX WP1: 2000-364475/31.

XX DR P-PSDB; B03121.

XX PT New immunogenic recombinant hybrid M protein comprising amino-terminal peptide fragments of streptococcal M protein useful as vaccine against rheumatic fever and infections leading to rheumatic fever

XX PS Disclosure; Fig 11A-C; 62pp; English.

XX CC The invention relates to multivalent immunogenic hybrid group A streptococcal M proteins comprising N-terminal peptide fragments of M proteins that elicit opsonic antibodies against multiple serotypes of group A streptococci (e.g., Streptococcus pyogenes). The antibodies generated using the hybrid proteins are against one or more M protein serotypes, where at least one of the serotypes is M1, M3, M18 or M19.

CC The invention also encompasses a recombinant DNA molecule comprising a nucleotide sequence that encodes a multivalent hybrid M protein; and a method for immunising a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic heart disease. In a patient with ARF, antibodies formed during a group A streptococcal infection are also cross-reactive with heart tissue, which indicates that the streptococci and host tissue contain similar antigenic motifs. The new multivalent vaccines are capable of raising sero-specific antibodies against various serotypes of group A streptococci which are not cross-reactive with human heart tissue. Sequences A57893-A57902 represent DNAs encoding multivalent hybrid streptococcus pyogenes M proteins generated in the disclosure of the invention.

XX Sequence 1029 BP; 424 A; 168 C; 226 G; 211 T; 0 other;

Query Match 26.1%; Score 301.8; DB 21; Length 1029;

Best Local Similarity 68.3%; Pred. No. 3.1e-57;

Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;

QY 486 gagagtcgttatactaggtacgacagagaataagctaaataattattgacgactct 545

Db 3 gagagtcgttatactaggtacgacagagaataagctaaataattattgacgactct 62

QY 546 tgacgcaaaagaacatgaattacaacacagaaatgagaagttatctctgcagaagcgtga 605

Db 63 tgacgcaaaagaacatgaattacaacacagaaatgagaagttatctctgcagaagcgtgt 122

QY 606 tggtaatccttaggggaagttatagaagatcttcgacacaacaatcccgcaatacaaatat 665

Db 123 tcttaggggacggtgagaaacccgg-----acaaagc 155

QY 666 acgtttacgtcacgaaacacagagacttaaaagcgagattagagaatgcgaatggaagtgc 725

Db 156 acgagaactcttacaagtagatgacgtagagaactctatgttacaagtaataatgacaa 215

QY 726 aggaagagatttttaagagagcgtgactctgttagatcaggttacacaattatatactaa 785

Db 216 cttagtcgacgcggtgactaggggtacataataatgacccgcaaaagcgcaaaagagactct 275

QY 786 acataatagtaattacacaataataatgcacaagctggcagacttgacctgagacaaaa 845

Db 276 tgacaagtagatgactagaaaacccatgacttaaaactaagaatgaaagggttaaaaaactga 335

QY 846 ggctgaatatctaaagcccttaattgattgggtgagaggctgttacaagagttaaatat 905

Db 336 gaatgaagggttaaaaaactgagaatgaagggttaaaaaactgagaatgaagggttcaaaaac 395

QY 906 cga-----tgfcgcgactagctcagacagatactctggaagaaagtaacaagaactgc 959

Db 396 tgagccatgggttcgcgactagctcagacagatactctggaagaaagtaacaagaactgc 455

QY 960 tgacaagtttgagatagaagaacaatcagtttaaaacttaagaatgaagtagacttaatttaa 1019

Db 456 tgacaagtttgagatagaagaacaatcagtttaaaacttaagaatgaagtagacttaatttaa 515

QY 1020 taataaagcgttaaaagatcataatgatgagttaaactgaagtgtagtaatgcttaaga 1079

Db 516 taataaagcgttaaaagatcataatgatgagttaaactgaagtgtagtaatgcttaaga 575

QY 1080 gaaactacgtataaaatgataaaatcactatctgaaaaagctagtaaaatcaagaattaga 1139

Db 576 gaaactacgtataaaatgataaaatcactatctgaaaaagctagtaaaatcaagaattaga 635

QY 1140 ggcacgttaagtaaaagctt 1158

Db 636 ggcacgtaaggctgatctt 654

RESULT 10

ID Q45214 standard; DNA; 861 BP.

XX AC Q45214;

XX DT 02-NOV-1994 (first entry)

XX DE Recombinant M24-M5-M6 gene.

XX KW Primer: PCR; amplify; polymerase chain reaction; construct; hybrid;
M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
tandem; PK223.3; protective epitope; tissue-cross-reactive epitope;
restriction enzyme site; multivalent M protein; immunisation; group A;
streptococci; rheumatic fever; rheumatic heart disease; group A;
antibody; heart tissue; antigen; serotype; mucosal; ss.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT m15c_difference 781..783

XX FT /tag= a

XX FT /codon= seq:gga, aa:Arg

XX PN W09406421-A.

XX PD 31-MAR-1994.

XX PF 15-SEP-1993; 93MO-US08703.

XX PR 16-SEP-1992; 92US-0945954.

XX PA (UYTE-) UNIV TENNESSEE RES CORP.

XX PI Dale JB, Lederer JW;

XX DR WPI; 1994-118122/14.

XX DR P-PSDB; R50992.

XX New immunogenic hybrid proteins derives from streptococcal M proteins
PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes

XX PS Disclosure; Fig 1; 67pp; English.

XX CC The sequences given in Q45214-23 encode hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These tetraivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC PK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
CC to an immunogenic carrier.

XX SQ Sequence 861 BP; 331 A; 147 C; 199 G; 184 T; 0 other;

Query Match

Best Local Similarity 20.8%; Score 241; DB 15; Length 861;

Matches 244; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 910 gtcgcgactaggctcagacagataactctggaaaaagtcacaagaacgtgctgacaagt 969
|||||
Db 4 gtcgcgactaggctcagacagataactctggaaaaagtcacaagaacgtgctgacaagt 63
QY 970 gagatagaaaaacaatacgtttaaaacttaagaatagtgacttaagtttttaataaagcg 1029
|||||
Db 64 gagatagaaaaacaatacgtttaaaacttaagaatagtgacttaagtttttaataaagcg 123
|||||
QY 1030 ttaaaagatcataatgatgatgagtttaactgaaggttgagtaataatgctaaaggaactacgt 1089
|||||
Db 124 ttaaaagatcataatgatgatgagtttaactgaaggttgagtaataatgctaaaggaactacgt 183
QY 1090 aaaaatataaaatcactatctgaaaaagctgataaaattcaagaattagaggcacgttaag 1149
|||||
Db 184 aaaaatataaaatcactatctgaaaaagctgataaaattcaagaattagaggcacgttaag 243
QY 1150 taaaagctt 1158
| |||
Db 244 gctgatctt 252
RESULT 11
A57893
ID A57893 standard; DNA; 861 BP.
XX AC A57893;
XX DT 10-OCT-2000 (first entry)
XX DE S. pyogenes hybrid M protein (M24-M5-M6) DNA, SEQ ID NO:1.
XX KW Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARP;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia; ds.
XX OS Streptococcus pyogenes.
XX OS Synthetic.
XX PN US6063386-A.
XX PD 16-MAY-2000.
XX PF 15-SEP-1997; 97US-0937271.
XX PR 16-SEP-1992; 92US-0945954.
XX PA (UYTE-) UNIV TENNESSEE RES CORP.
XX PI Lederer JW, Dale JB;
XX DR WPI; 2000-364475/31.
XX DR P-PSDB; B03113.
XX PT New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX PS Disclosure; Fig 1A-B; 62pp; English.
XX CC The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections; and a
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting

Query Match	20.7%	Score 239.4	DB 21	Length 522
Best Local Similarity	97.6%	Pred. No. 1.le-43		
Matches 243	Conservative 0	Mismatches 6	Indels 0	Gaps 0
QY 910	gtcgcgactaggtctcagacagatactctggaagagtcacagaacgtgctgacaaagttt	969		
Db				
4	gtcgcgactaggtctcagacagatactctggaagagtcacagaacgtgctgacaaagttt	63		
QY 970	agatagagaacaatacagtttaaaacttaagaatagtgacttaagttttaataataaacgcg	1029		
Db				
64	gagacagagaacaatacagtttaaaacttaagaatagtgacttaagttttaataataaacgcg	123		
QY 1030	ttaaaagatcataatgatgatgttaactggaagagttgagtaatgctaaagagaaactacgt	1089		
Db				
124	ttaaaagatcataatgatgatgttaactggaagagttgagtaatgctaaagagaaactacgt	183		
QY 1090	aaaatgataaaatcacatctctgaaaaagctagttaaattcaagatttagggcagctaa	1149		
Db				
184	aaaatgataaaatcacatctctgaaaaagctagttaaattcaagatttagggcagctaa	243		

Query Match	10.7%;	Score 124.4;	DB 15;	Length 561;
Best Local Similarity	55.5%;	Pred. No. 1.5e-18;		
Matches 295;	Conservative	0;	Mismatches 216;	Indels 21; Gaps
7	atgtctgcgactaggtrctcagacagatactcttgaaaaagtacaaagaacgtgtgacaag	66		

Db 1 atgtcgcgactaggtctcagacagatactcttgaaaaagtaacaaagatcgcgactagg 60
Qy 67 tttagatagaaaaacatactgtaaaactaagaatagtagtacttaagttaataataa 126
Db 61 tctcagacagatactctgaaaaagtaacaaagatcgcgactaggctcagacagatact 120
Qy 127 gcgttaaaagatcataatgatgatgtaaaactgaagagtgtagtaagtctaagaagaaacta 186
Db 121 ctggaaaaagtaacaaagatcgcgcgtgactagggttacaataaataatgacccgcaaga 180
Qy 187 cgtggatccgcgtgactagggttacaataaataatgacccgcaagaagaaactctt 246
Db 181 gcaaaagacgcgtgactagggttacaataaataatgacccgcaagaagaaactcgtg 240
Qy 247 gacagtgatgactagaaaaactgacttaaaactaagaatgaaggtttaaaaactgsg 306
Db 241 actagggttacaataaataatgacccgcaagaagaaactgacagagtggttctcagg 300
Qy 307 aatgaaggtttaaaaactgagaatgaaggtttaaaaactgagaatgaaggtttaaaaact 366
Db 301 gggacggtagaaaaacccgcaagaagacgaagagtggttctcagggttctcagg 360
Qy 367 gaggtcgac-----agagtggttctcagggtgacggtgtagaaaaacccgcaagaacga 420
Db 361 ccggacaaagcagcagagagtggttctcagggtgacggtgtagaaaaacccgcaagaacga 420
Qy 421 gaacttcttaacaagatgacgtgagagaaactctatgtttacaagctaataatgacaagtta 480
Db 421 cca-----tggagagtgcttatactaggcatcgcgcagagaataagcta 465
Qy 481 ccatggagagtgcttatactaggcatcgcgcagagaataagctaataaaaaa 532
Db 466 aaaaaagagtgcttatactaggcatcgcgcagagaataagctaataaaaaa 517

RESULT 15
A57896
ID A57896 standard; DNA; 561 BP.
AC A57896;
DT 10-OCT-2000 (first entry)
DE S. pyogenes hybrid M protein (M24-M5-M6-M19) DNA, SEQ ID NO:7.
XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia; ds.
XX Streptococcus pyogenes.
OS Synthetic.
XX US0603386-A.
PN 16-MAY-2000.
XX 15-SEP-1997; 97US-0937271.
PF 16-SEP-1992; 92US-0945954.
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX Lederer JW, Dale JB;
XX WPI; 2000-364475/31.
DR P-PSDB; B03116.
XX New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX

PS Disclosure; Fig 7A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A streptococcal M proteins comprising N-terminal peptide fragments of M proteins that elicit opsonic antibodies against multiple serotypes of group A streptococci (e.g., Streptococcus pyogenes). The antibodies generated using the hybrid proteins are against one or more M protein serotypes, where at least one of the serotypes is M1, M3, M18 or M19. The invention also encompasses a recombinant DNA molecule comprising a nucleotide sequence that encodes a multivalent hybrid M protein; and a method for immunising a mammal against streptococcal infections, comprising administering an immunogenic multivalent hybrid M protein to the mammal. The multivalent hybrid M proteins are useful for eliciting opsonic or protective antibodies to the M proteins of different serotypes of group A streptococci, and may therefore be used as vaccines to protect against and control infection by type A streptococci. Type A streptococci are not only responsible for streptococcal pharyngitis (strep throat), forms of pneumonia and a condition resembling toxic shock, but are also involved in the development of acute rheumatic fever (ARF) and rheumatic heart disease. In a patient with ARF, antibodies formed during a group A streptococcal infection are also cross-reactive with heart tissue, which indicates that the streptococci and host tissue contain similar antigenic motifs. The new multivalent vaccines are capable of raising sero-specific antibodies against various serotypes of group A streptococci which are not cross-reactive with human heart tissue. Sequences A57893-A57902 represent DNAs encoding multivalent hybrid Streptococcus pyogenes M proteins generated in the disclosure of the invention.

XX Sequence 561 BP; 217 A; 108 C; 151 G; 85 T; 0 other;

Query Match 10.7%; Score 124.4; DB 21; Length 561;
Best Local Similarity 55.5%; Pred. No. 1.5e-18;
Matches 295; Conservative 0; Mismatches 216; Indels 21; Gaps 2;

Qy 7 atgttcgcgactaggtctcagacagatactctgaaaaagtaacaaagatcgcgcaag 66
Db 1 atgttcgcgactaggtctcagacagatactctgaaaaagtaacaaagatcgcgactagg 60
Qy 67 tttagatagaaaaacatactgtaaaactaagaatagtagtacttaagtttaataataa 126
Db 61 tctcagacagatactctgaaaaagtaacaaagatcgcgactagggtctcagacagatact 120
Qy 127 gcgttaaaagatcataatgatgatgtaaaactgaagagtgtagtaagtctaagaagaaacta 186
Db 121 ctggaaaaagtaacaaagatcgcgcgtgactagggttacaataaataatgacccgcaaga 180
Qy 187 cgtggatccgcgtgactagggttacaataaataatgacccgcaagaagcaagaagctctt 246
Db 181 gcaaaagacgcgtgactagggttacaataaataatgacccgcaagaagcaagaagccgtg 240
Qy 247 gacaagtgatgactagaaaaacccatgacttaaaactaagaatgaaggtttaaaaactgag 306
Db 241 actagggttacaataaataatgacccgcaagaagcaagaagatgacagagtggttctcagg 300
Qy 307 aatgaaggtttaaaaactgagaatgaaggtttaaaaactgagaatgaaggtttaaaaact 366
Db 301 gggacggtagaaaaacccgcaagaagcagcaagagtggttctcagggtgacggttaga 360
Qy 367 gaggtcgac-----agagtggttctcagggtgacggttagaataacccgcaagaacga 420
Db 361 ccggacaaagcagcagagagtggttctcagggtgacggttagaataacccgcaagaacga 420
Qy 421 gaacttcttaacaagatgacgtagagaactctatgtttacaagctaataatgacaagtta 480
Db 421 cca-----tggagagtgcttatactaggcatcgcgcagagaataagcta 465
Qy 481 ccatggagagtgcttatactaggcatcgcgcagagaataagctaataaaaaa 532
Db 466 aaaaaagagtgcttatactaggcatcgcgcagagaataagctaataaaaaa 517

Tue May 15 07:27:08 2001

Search completed: May 13, 2001, 04:57:44
Job time: 8144 sec

us-09-151-409-15.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 04:30:06 ; Search time 66.83 Seconds
(without alignments)
3025.436 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcatgcatgctgcgactag.....aggcacgtaagtaaaagctt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	419	36.2	741	3	US-08-937-271-3
2	328.2	28.3	822	3	US-08-937-271-5
3	301.8	26.1	741	3	US-08-937-271-12
4	301.8	26.1	1029	3	US-08-937-271-16
5	241	20.8	861	3	US-08-937-271-1
6	239.4	20.7	522	3	US-08-937-271-14
7	151	13.0	1332	2	US-08-937-271-14
8	124.4	10.7	561	3	US-08-937-271-7
9	82.6	7.1	7218	1	US-08-232-463-14
10	76.2	6.6	408	3	US-08-937-271-21
11	53.2	4.6	204	3	US-08-937-271-19
12	53.2	4.6	918	3	US-08-937-271-9
13	42.6	3.7	6243	2	US-09-056-075-1
14	42	3.6	2223	3	US-08-257-073-4
15	41.6	3.6	6755	3	US-08-931-999-4
16	41.4	3.6	1498	1	US-08-118-469A-1
17	41.4	3.6	1498	1	US-08-909-119-1
18	40.2	3.5	4766	5	PCT-US93-07261-10
19	39.8	3.4	5181	1	US-08-257-073-10
20	39.6	3.4	243	1	US-08-182-175A-56
21	39.6	3.4	243	1	US-08-474-633A-74
22	39.6	3.4	243	5	PCT-US92-06412-56
23	39.4	3.4	5361	4	US-08-973-462-2
24	39.4	3.4	6152	4	US-08-973-462-1
25	38.8	3.4	3279	4	US-08-446-137B-1
26	38.6	3.3	1519	1	US-07-971-759-19
27	38.6	3.3	2447	2	US-09-014-969-14

28 38.6 3.3 3645 2 US-08-663-112-1
29 38.4 3.3 453 3 US-08-714-918-9
30 38.4 3.3 453 4 US-09-265-315-9
31 38.4 3.3 453 4 US-09-265-315-9
32 38.2 3.3 2258 1 US-07-720-589-1
33 38.2 3.3 2258 2 US-08-785-190-1
34 38.2 3.3 2258 5 PCT-US92-05539-1
35 38.2 3.3 2672 4 US-09-214-564A-5
36 38.2 3.3 2815 4 US-09-214-564A-1
37 38.2 3.3 3143 1 US-08-485-621-1
38 38.2 3.3 3143 2 US-08-973-831-1
39 38.2 3.3 3143 5 PCT-US96-09530A-1
40 38 3.3 39 3 US-08-937-271-26
41 38 3.3 47 3 US-08-937-271-28
42 38 3.3 1320 1 US-08-257-073-15
43 37.8 3.3 5433 4 US-08-929-329-1
44 37.8 3.3 19124 2 US-08-487-826B-13
45 37.6 3.2 1560 1 US-07-813-584A-1

ALIGNMENTS

RESULT 1
US-08-937-271-3
; Sequence 3, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 741 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
US-08-937-271-3

Query Match 36.2%; Score 419; DB 3; Length 741;
Best Local Similarity 78.9%; Pred. No. 1.4e-98;
Matches 585; Conservative 0; Mismatches 0; Indels 156; Gaps 1;

QY 7 atggtcgacgaagtcagacagatctctgaaaaagtcacaaagcgtgctgacaag 66
DB 1 ATGGTCCGACGTAGTCTCAGACAGATCTCTGAAAAAGTACAAAGCTGCTGACAAG 60

QY 67 ttgagatagaaaaaatacgttaaaacttaagaatagtagtacttaagtttaataataa 126
DB 61 TTTGAGATAGAAACATACCTTTAAACTTTAAGAATAGTGACCTAACTTTAATAATAA 120

QY 127 gcgttaaaagatcataatgtagtactgaagagtgtagtgaatgtaagaagaacta 186
DB 121 GCGTTAAAAAGATCATATATGATGAGTAACTGAAAGAGTGAATGCTAAAGAGAACA 180

QY 187 cgt----- 189
DB 181 CGTAAATGATAAATCACTATCTGAAAAAGCTAGTAAATCAAGAAATTAGAGCAGCT 240

QY 190 ----- 189
DB 241 AAGGCTGATCTTTGAAAAAGCATTAGAGGCGCAATGAATTTTCAACAGCGGATTCAGCT 300

QY 190 ----- 210
DB 301 AAATCAAAACCTTAGACGACAGAGAAAGCTGATCTTGAGGATCCGCGTGACTAGGGGT 360

QY 211 acaataaatgacccgcacaaagagcaaaagagctcttgacaagtagtagtgaataaaccat 270
DB 361 ACATAAATGACCCGCAAGAGAGCAAAAGAGCTCTTGACAAGTATGAGCTAGAAACCAT 420

QY 271 gacttaaaactaaagaatgaaggggttaaaactgagaatgaaggttaaaactgagaat 330
DB 421 GACTTAAAACTAAGAATGAAGGGTTAAAACTGAGATGAAGGGTTAAAACTGAGAA 480

QY 331 gaaggggttaaaactgagaatgaaggggttaaaactgaggtcgacagagtggttccttagg 390
DB 481 GAAGGGTTAAAACTGAGAAATGAGGGTTAAAACTGAGGTCGACAGAGTGTTCCTTAGG 540

QY 391 gggacgggtgaaacccgcacaaagacagagaaacttcttaacaaagtatgacgtagagaac 450
DB 541 GGGACGGTGAAGAACCCGACAAAGACAGAACTTCTTAACAAAGTATGACGTAGAGAAC 600

QY 451 tctatgttaacagctaaatgaacaggttaccatggaggtgaggtgcgttatactaggcatacg 510
DB 601 TCTATGTACAACTATATGACAAAGTTACCATGAGAGTGCCTTATACCTAGGCATACG 660

QY 511 ccagaagataagctaaaaaaattatgacgactcttgacgcgaagaacatgaattacaa 570
DB 661 CCAGAAGATAGCTAAAAAAAATTATTGACGATCTTGACGCAAAAGAACATGAATTACAA 720

QY 571 caacagaatgagaagtattct 591
DB 721 CAACAGAATGAGAAGTTATCT 741

RESULT 2
US-08-937-271-5
; Sequence 5, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA

ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..822
US-08-937-271-5

Query Match 28.3%; Score 328.2; DB 3; Length 822;
Best Local Similarity 87.6%; Pred. No. 2.5e-75;
Matches 404; Conservative 0; Mismatches 3; Indels 54; Gaps 2;

QY 185 tacgttgatcccgctgactaggggtacaaataatgacccgcacaaagagcaaaagagctc 244
DB 362 TTCCAGGATCCCGCTGACTAGGGGTACATAAATGACCCCAAGAGCAAAAGAGCTC 421

QY 245 ttgacaagtatgactagaaaaccatgacttaaaactaaagaatgaaggggttaaaaaactg 304
DB 422 TTGACAAGTATGAGCTAGAAAACCATGACTTAAAAACTAAGAATGAAGGGTTAAAAACTG 481

QY 305 aqaatgaaggggttaaaaaactgagaatgaaggggttaaaaaactgagaatgaaggggttaaaaa 364
DB 482 AGAATGAAGGGTTAAAAACTGAGATGAAGGGTTAAAAACTGAGAAATGAAGGGTTAAAAA 541

QY 365 ctgag-----gtcgacagagtggttctctagggggacgg 397
DB 542 CTGAGGTGACCCAGGAAACCCAGGCTGTCCAGTCGACAGAGTGTTCTTAGGGGACGG 601

QY 398 tagaaacccgcgacaaagcacgagaactctttaaagaatgacgtagagaactctatgt 457
DB 602 TAGAAAACCCGGACAAAGCACGAGAACTTCTTAACAAGTATGACGTAGAGAACTCTATGT 661

QY 458 tacaagctaaatgacaaggttacc-----atgagagag 490
DB 662 TACAAGCTAATAATGACAAGTTACCATGGCCAGGAAACCCAGCTGTTCCACCATGAGAG 721

QY 491 tgcgttatactaggcatagccagaagataaagctaaataaaataattattgacgactctgacg 550
DB 722 TGCCTTATAGTATAGCATACGCCAGAGATAAGCTAAATAAATAATTATTGACGATCTTACG 781

QY 551 caaaagaacatgaattacaaacacagaatgagaagtattct 591
DB 782 CAAAAGAACATGAATTACAAACACAGAATGAGAAGTTATCT 822

Sequence 12, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..741
US-08-937-271-12

Query Match 26.1%; Score 301.8; DB 3; Length 741;
Best Local Similarity 68.3%; Pred. No. 1.4e-68;
Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;
QY 486 gagagtgcttactagcgcagacgagataagctaaataattattgacgatct 545
|||||
3 GAGAGTCGTTACTAGGATAGCGCAGAGATAGCTAAAAAATAATTATGACGATCT 62
QY 546 tgacgaaaagacatgaattacaacagagatgagagttatctctgcagacggtga 605
|||||
63 TGACGCAAAAGACATGAATTAACAACAGAGATGAGAGTTATCTGGATCCAGAGTGT 122
QY 606 tggtaacctaggaagattagaagattcttgacgcaaacatcccgcaatacaaaatat 665
|||||
123 TCCTAGGGGACGGTAGAANACCGG-----ACAAAGC 155
QY 666 acgtttactgcagaaaaaagacttaaaagcgagattagagaatgcaatggaagtgc 725
|||||
156 ACGAAGACTTCTTAACAAGTATGACGTAGAACCTATGTTACAGGCTTAATAATGACAA 215
QY 726 aggaagagatttaagagagctgggtacctgttgatcaggtttacacattatatactaa 785
|||||
216 CTTAGTCGCGCGTACAGTGGGGTACAATAATGACCGCAAGAGCAAGAGAGCTCT 275
QY 786 acataatagtaattaccacaataataatgcacaaagctggcagacttgacctgagacaaa 845

Db 276 TGACAAGTATGAGCTAGAAAACCATGACTTAAAAAATAAGAGGTAAAAAAGTGA 335
QY 846 ggctgaatatctaaaagcccttaattgattggcgtgagagctgtttacaagagttaaatat 905
|||||
Db 336 GAATCAAGGGTTAAAAAATAAGAGGTAAAAAATAAGAGGTAAAAAAGTGA 395
QY 906 cga-----tgtcgactaggtctcagacagatactctggaataagctaaagacgtgc 959
|||||
Db 396 TGAGCCATGGTGCAGTACTAGTCTCAGACAGATACTCTGAAAAAGTACAGAACGTGC 455
QY 960 tgacaagtgtgagatagaacacatactttaaaacttaagaacttaagaacttaagtttaa 1019
|||||
Db 456 TGACAAGTTCAGATAGAAAAACATACTGTTAAAACTTAAAGATAGTGAAGTTTAA 515
QY 1020 taataagcgttaaaagatcataatgatgatgaattaaactgaagagttgagtaactgaaga 1079
|||||
Db 516 TAATAAAGCGTTAAAGATCATATGATGATGATTAACCTGAAGAGTTGAGTATGCTTAAGA 575
QY 1080 gaaactacgtaaaaatgataaaatacactatctgaaaaagctagtaaaattcaagaattaga 1139
|||||
Db 576 GAACTACGTAAAAATGATAAATCACTATCTGAAAAAGCTAGTAAAAAATCAAGAATTAGA 635
QY 1140 ggcacgtgaagtaaaagcgtt 1158
|||||
Db 636 GGCACGTAAAGCTGATCTT 654

RESULT 4
US-08-937-271-16
Sequence 16, Application US/08937271
Patent No. 6063386
GENERAL INFORMATION:
APPLICANT: Dale, James B.
APPLICANT: Lederer, James W.
TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1029 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:

NAME/KEY: CDS
LOCATION: 1..1029
US-08-937-271-16

Query Match 26.1%; Score 301.8; DB 3; Length 1029;
Best Local Similarity 68.3%; Pred. No. 1.5e-68;
Matches 464; Conservative 0; Mismatches 182; Indels 33; Gaps 2;

QY 486 gagagtgcttatactaggcctacgcagagataagcttaaaataattattgacgatct 545
DB 3 GAGAGTGGCTTACTAGGCATACGCCAGAGATAAGCTAAAAAATAATTATGACGATCT 62

QY 546 tgacccaaagacatgaattacaaacagagatgaagaattctctcgcagacggtga 605
DB 63 TGACGCAAAAGAACATGAATTTACAAACACAGAAATGAGAAGTTATCTGGATCCAGAGTGT 122

QY 606 tggtaactcctagggaagttatagaagatcttgcagcaacaatcccgcaatacaaatat 665
DB 123 TCCTAGGGGCGGTAGAAACCCGG-----ACAAAGC 155

QY 666 acgttttacgtcacgaaacagagacttaaaagcgagattgagagatgaatggaagtgc 725
DB 156 ACGAGAACTTCTTAACAAGTATGAGTAGAGAACTCTATGTTACAGCTAATAATGACAA 215

QY 726 aggaagagatttaagagagctggttaccttgttagatcaggttacacaattatatactaa 785
DB 216 CTTAGTCGACCGCTGACTAGGGGTGACAAATTAATGACCCGCAAGACAAAGAGCTCT 275

QY 786 acataatagtaattaccacaataataatgcacaagctggcagacttgacctggagacaaa 845
DB 276 TGACAAAGTATGAGCTAGAAACCATGACTTAAAACTAAGAATGAAGGTTAAAAACTGA 335

QY 846 ggcgaatcttaaaagccttaagtattggtgagagcgtgttacagagtttaaatat 905
DB 336 GAATGAAGGTTAAAACTGAGAATGAAGGGTTAAAACTGAGAACTGAGAGTTAAAAAC 395

QY 906 cga-----tgtcgactaggtctcagacagatactctggaaaaagtcagaacgtgc 959
DB 396 TGAGCCATGGTTCGGACTAGTCTCAGACAGATACTCTGGAAAAAGTACAAAGACGTGC 455

QY 960 tgacaagtttgagatgaagaaacaaatcgttaaaacttaagaatagtgacttaagtttaa 1019
DB 456 TGACAAGTTTGAGATGAGAAACAAATACGTAAAACTTAAGAACTAGTACTTAAGTTTAA 515

QY 1020 taataagcgttaaaagataatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 1079
DB 516 TATAAAGCGTTAAAGATCATATGATGAGTTAACTGAAGAGTTGAGTAAATGCTAAAGA 575

QY 1080 gaaactacgttaaaatgataatcactatctgaaaaagcctagcaaaatcaagaattaga 1139
DB 576 GAAACTACGTAAAAATGATAATCATACTATCTGAAAAAGCTAGTAAAAATCAAGATTAGA 635

QY 1140 ggaacgttaagtaaaagcctt 1158
DB 636 GGCACGTAAGGCTGATCTT 654

RESULT 5
US-08-937-271-1
; Sequence 1, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 861 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
FEATURE:
NAME/KEY: CDS
LOCATION: 1..861
US-08-937-271-1

Query Match 20.8%; Score 241; DB 3; Length 861;
Best Local Similarity 98.0%; Pred. No. 5.2e-53;
Matches 244; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 910 gtgcgactaggtctcagacagatactctgaaaaagtcagaacgtgctgacaaattt 969
DB 4 GTCCGACTAGGTCTCAGACAGATACTCTGAAAAAGTACAAAGACGTGCTGACAAATT 63

QY 970 gagatagaaacaatacgtttaaacttaagaatagtgacttaagttttaataataaagcg 1029
DB 64 GAGATAGAAAAACAACTAGTTAAACTTAAGATAGTACTTAAGTTTAAATAATAAGCG 123

QY 1030 ttaaaagatcataatgatgatgagttgaagagttgagtaatgctaaagagaactacgt 1089
DB 124 TTAAGAGATCATAATGATGATGATTAAGTAACTGAAGAGTTGAGTAAATGCTAAAGAACTACGT 183

QY 1090 aaaaatgataaatcactatctgaaaaagcctagtaaaattcaagaattagagcgtaag 1149
DB 184 AAAAATGATAAATCACTATCTGAAAAAGCTAGTAAANTCAAGAAATTAGAGCGCTAAG 243

QY 1150 taaaagcctt 1158
DB 244 GCTGATCTT 252

RESULT 6
US-08-937-271-14
; Sequence 14, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue


```

Query Match      13.0%; Score 151; DB 2; Length 1332;
Best Local Similarity 100.0%; Pred. No. 6,4e-30;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps
QY 598 aacgggtgatggtaactcctagggaagtattagaagatcttgcagcaacaatcccgcaata 657
      |||
Db 1 AACGGTGATGGTAAATCTCTAGGGAAGTTATAGAAGATCTTGCAGCAAAACAATCCGCAATA 60
QY 658 caaatatatacgttttacctgcacgaataaacaggacttaaaagcgagatagagaatgcatg 717
      |||
Db 61 CAAATATACGTTTACGTCACGAAACAAGGACTTAAAGCGAGNTTAGAGATGCAATG 120
QY 718 gaagttgcaggaagagatatttaagagagctg 748
      |||
Db 121 GAAGTTGCAGGAAGAGATTTTAAAGAGAGCTG 151

RESULT 8
US-08-937-271-7
; Sequence 7, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/937,271
;; FILING DATE: 15-SEP-1997
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rosenman, Stephen J.
;; REGISTRATION NUMBER: 43,058
;; REFERENCE/DOCKET NUMBER: 481112.405C1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 561 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Streptococcus pyogenes
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..561
;; US-08-937-271-7

Query Match 10.7%; Score 124.4; DB 3; Length 561;
Best Local Similarity 55.5%; Pred. No. 3e-23;
Matches 295; Conservative 0; Mismatches 216; Indels 21; Gaps 2;
QY 7 atgtgcgactggttcacagacagatactctggaagaaagtaagaacgctgacaa 66
Db 1 ATGTGCGGACTAGGTCTCAGACAGATACCTCTGGAAAAAGTACAAGAGTCGCGACTAGG 60
QY 67 tttagatagaaacataatcgctaaacttaagaatagtgacttaagtttaataataaa 126
Db 61 TCTCAGACATACCTCTGGAAAAAGTACAAGAGTCGCGACTAGGTCTCAGACAGATAC 120
QY 127 gcgttaaaagatcataatgatgatgttaactgaagagttgagtaactgaagaaacta 186
Db 121 CTGAAAAAAGTACAAAGAGTACCGCGTGCACCTAGGAGTACATAAATGACCCGCAAGA 180
QY 187 cgtggtccgctgactgaggtgacataataatgacccgcaagagcaaaagagctctt 246
Db 181 GCAAAAGAACCGGTGACTAGGGTACANTAAATGACCCGCAAGAGCAAAAGAACCGGTG 240
QY 247 gacagtatgagctagaaaaaccatgacttaaaacttaagaatgaagggttaaaactgag 306
Db 241 ACTAGGGGTACAATAATGACCCGCAAGAGCAAAAGAGTCGACAGAGTGTTCCTAGG 300
QY 307 aatgaaggggttaaaactgagaatgaagggttaaaactgagatgaagggttaaaact 366
Db 301 GGGAGGTAGAAAACCGGACAAAGAGTGTTCCTAGGGGGACGGTAGAAGAC 360
QY 367 gaggctgcac-----agagttcttcctaggggacgttagaaaacccgacaaagcaga 420
Db 361 CCGGACAAAGCACGAGAGAGTGTTCCTAGGGGGACGGTAGAAGACCGGACAAAGCAGA 420
QY 421 gaactcttaacagatgacgtgagaaactctatgttaacagtaataatgacagttta 480
Db 421 CCA-----TGAGAGTGGCTTATACCTAGGCATACGCCAGAGATAGCTTA 465
QY 481 ccatggagagtgctgtatactaggcatcacgcagaaagataagcttaaaaaaa 532
Db 466 AAAAAAGAGTGGCTTATACCTAGGCATACGCCAGAGATAGCTTAAGCTTA 517

RESULT 9
US-08-232-463-14/c
; Sequence 14, Application US/08232463

;; Patent No. 5670367
;; GENERAL INFORMATION:
;; APPLICANT: DORNER, F.
;; APPLICANT: SCHEIFLINGER, F.
;; APPLICANT: FALKNER, F. G.
;; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;; NUMBER OF SEQUENCES: 52
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: 1800 Diagonal Road, Suite 500
;; CITY: Alexandria
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22313-0299
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/232,463
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/935,313
;; FILING DATE:
;; APPLICATION NUMBER: EP 91 114 300.6
;; FILING DATE: 26-AUG-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 30472/114 IMM
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703)836-9300
;; TELEFAX: (703)683-4109
;; TELEX: 899149
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7218 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; CLONE: pTZgpt-Fls
;; US-08-232-463-14

Query Match 7.1%; Score 82.6; DB 1; Length 7218;
Best Local Similarity 8.9%; Pred. No. 3.7e-12;
Matches 40; Conservative 239; Mismatches 168; Indels 0; Gaps 0;
QY 115 tttaataaaacgcttaaaagatcataatgatgagtgactgaagagtgagtaagct 174
Db 1511 TTTCAAAAACGGCATGTAGGCATCACTGTAATCTATCTATGCAAGTACTTAAGAG 1452
QY 175 aaagagaaactgctgacgcgcgtgactaggggtacataataatgacccgcaagagca 234
Db 1451 ATACAGAATTGTGTACRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1392
QY 235 aaagagactcttgacaagtatgactgagaaacacacactgacttaaaactagaagatga 294
Db 1391 RNR 1332
QY 295 ttaaaactgagaatgaagggttaaaactgagaatgaagggttaaaactagaagatga 354
Db 1331 RNR 1272
QY 355 ggggttaaaactgaggggtcagacagagtttccctaggggagcgttagaaaacccgga 414
Db 1271 RNR 1212
QY 415 gcacgagactcttcaacaagtatgacgtagagactctatgttaacagtaataatgac 474

[illegible]

```

RESULT 10
US-08-937-271-21
; Sequence 21, Application US/08937271
; Patent No. 606386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESS: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
;

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/VMS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
;

```

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 48112.403C1

```

;
; TELECOMMUNICATION INFORMATION
;
; TELEPHONE: (206) 622-4900
;
; TELEFAX: (206) 682-6031
;
; INFORMATION FOR SEQ ID NO: 21:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

ANIT-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1 405

```

LOCATION:
US-08-937-271-21

Query Match 6.6%; Score 76.2; DB 3; Length 408;
Best Local Similarity 71.5%; Pred. No. 5.8e-11;
Matches 118; Conservative 0; Mismatches 38; Indels 9; Gaps 1;

Qy 480 accatggagagtgcgttatactaggcatcgcgaagataaagctaaaaaattatga 539
 |||||
 Db 150 ACCATGGAGAGTGCCTTATACTAGGCATACGCCACCAAGATAAAGCTAAAAAATGCAGGA 209
 |||||

QY	540	cqatcttgacgcaaaagaacatgaattacaacaacagatgagaggttatctctgcagaa	599
Dδ	210	TGCTAGGAGTGTTAATGGACAGTTCCTCAGACATGTTAA-----ATTATCCATAA	260

ATTACCAIAA Z80

Qy 600 cggtgatgtaatccttaggaagttatagaagatcttcgacgaaa 644
|||
Db 261 CGGTGATGTTATCCCTAGGGAAGTTATAGAGATCTTCGACGAGA 305

RESULT 11

US-08-937-271-19
; Sequence 19, Application US/08937271
; Patent No. 6063386

```

:
:
: GENERAL INFORMATION:
:
: APPLICANT: Dale, James B.
: APPLICANT: Lederer, James W.
: TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
: TITLE OF INVENTION: VACCINE
:
: NUMBER OF SEQUENCES: 40
:
: CORRESPONDENCE ADDRESS:
: ADDRESS: SEED AND BERRY
: STREET: 6300 Columbia Center, 701 Fifth Avenue
:
: City: Seattle
:
: STATE: Washington
:
: COUNTRY: USA
:
: ZIP: 98104
:
:

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;

```

; APPLICATION NUMBER: US/08/937,271
 ; FILING DATE: 15-SEP-1997
 ; CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1

TELECOMMUNICATION INFORMATION
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

```

; SEQUENCE: (200) 002 0001
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 204 base pairs

```

;
; strand: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

```

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes

```

```

; ORGANISM: streptococcus pyogenes
;
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..201

```

US-08-937-271-19

7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844

Query Match 4.6%; Score 53.2; DB 3; Length 204;
Best Local Similarity 87.9%; Pred. No. 3.5e-05;
Matches 58; Conservative 0; Mismatches 8; Indels

Qy 176 aagagaaactacgtggtccgccggtgactagggggtacaataaatgacccgcgaagagcaa 235
Db 35 AAAAGTACAAGAAGCATCCGCCGTACTAGGGGTACAATAATGACCCGCAAGAGCAA 94

Qy 236 aagaag 241

Db 95 AAGAAG 100

RESULT 12

US-08-937-271-9
; Sequence 9, Application US/08937271
; Patent No. 6063386

; GENERAL INFORMATION:

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-073-4

Query Match 3.6%; Score 42; DB 1; Length 2223;
Best Local Similarity 46.3%; Pred. No. 0.06; Indels 0; Gaps 0;
Matches 138; Conservative 0; Mismatches 160;

```
QY 83 atactttaaacttaagaatagtgatgtaatttaataataaagcggttaaaagatcata 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1892 AAATATTTCAACAAAATATTCTAGAAAATGATGTTCTTAATCAAGAAACGGAGGAAGAA 1951

QY 143 atgatgagtgtaactgaagagtgatgtaactgaagaagaaactacgtggatccgccgtga 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1952 TGGAAAACACAGTTGACCAATCCCAAGCAAAATAGAAAGTGAAGTGGATGCCCTCGCAC 2011

QY 203 ctgaggggtacataaatgaccgcgaagagcaaaagcgtaaaagcgctcttgacaagtatgagctag 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2012 CAAAATAATAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2071

QY 263 aaacacatgacttaaaactaagaatgaagggttttaaactgagaatgaagggttataaaa 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2072 AAAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131

QY 323 ctgagaatgaagggttataaaactgagaatgaagggtttataaaactgaggtcgacagagt 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2132 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATAGT 2189
```

RESULT 15
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: UT0007
; US-08-931-999-4

Query Match 3.6%; Score 41.6; DB 3; Length 6755;
Best Local Similarity 47.5%; Pred. No. 0.11;
Matches 134; Conservative 0; Mismatches 169; Indels 1; Gaps 1;

```
QY 42 aaagtacaaagacgtgctgacaaagtttgagatagaaacaatacgtttaaacttaagaa 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6112 AACCGAAAAAGACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6171

QY 102 tagtgacttaagttttaataataaagcgtttaaagatcataatgatgatgactgaaga 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6172 AACGGGACAGACGACCAAGAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6231

QY 162 gttgagtaagtctaaagagaaactacgtgactccgcgtagtaggggtacataataatga 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6232 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6291

QY 222 ccgcgaagagcaaaagaagctcttga-caagtatgagctagaacacccatgactcttaaaa 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6292 AACCCGCAAGAGCGAGCCCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6351

QY 281 ctaagaatgaagggtttataaaactgagaatgaagggtttataaaactgagaatgaagggttaa 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6352 CCAAAACCAAAAGGAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6411

QY 341 aaactgagaatgaagggttaaaaa 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6412 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6435
```

Search completed: May 13, 2001, 04:59:08
Job time: 1742 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2001, 01:57:50 ; Search time 1036.68 Seconds
(without alignments)
9758.461 Million cell updates/sec

Title: US-09-151-409-15
Perfect score: 1158
Sequence: 1 gcagcagtcgcgactag.....aggcacgtaagtaaaagcgtt 1158

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*
44: gb_est44:*
45: gb_est45:*
46: gb_est46:*
47: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query			DB	ID	Description
			%	Match	Length			
C	1	74.8	6.5	997	229	CNS005TE	AL060767	Drosophila
C	2	72.4	6.3	839	229	CNS0122R	AL101037	Drosophila
C	3	71.8	6.2	1447	146	BF273312	BF273312	GA_EB001
C	4	70.6	6.1	1885	166	BE420745	BE420745	HMM002.B0
C	5	69	6.0	1282	146	BF264918	BF264918	HV_CEA001
C	6	68	5.9	1036	231	CNS03LWJ	AL250012	Tetraodon
C	7	67.2	5.8	106	146	BF264948	BF264948	HV_CEA001
C	8	66.4	5.7	1223	227	BI2981	BI2981	T24D11-Sp6
C	9	66	5.7	907	230	CNS021J4	AL176953	Tetraodon
C	10	65.4	5.6	1101	229	CNS0039G	AL063921	Drosophila
C	11	65	5.6	1101	229	CNS017RP	AL108415	Drosophila
C	12	64.8	5.6	853	223	AZ545275	AZ545275	ENTGB15TF
C	13	64.8	5.6	899	223	AZ533253	AZ533253	ENTBI02TF
C	14	64.6	5.6	1036	229	CNS00599	AL057797	Drosophila
C	15	64.2	5.5	1101	229	CNS0182P	AL108811	Drosophila
C	16	64.2	5.5	1135	230	CNS033GQ	AL226115	Tetraodon
C	17	63.8	5.5	1101	229	CNS0039L	AL063926	Drosophila
C	18	63.8	5.5	1184	227	BI3117	BI3117	T9K3-Sp6.2

Qy	1136	tagaggcagctaaagtaaaaagctt	1158
Db	28	TNNNNNNNNNNNNNTATNWT	6
RESULT	2		
CNS0122R		839 bp	DNA
LOCUS			GSS
DEFINITION		Drosophila melanogaster genome survey sequence SP6 end of BAC BACN07E20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION	AL101037		
VERSION	ALI01037.1	GI:5612648	
KEYWORDS			GSS.
SOURCE			fruit fly.
ORGANISM			Plasmid Drosophila melanogaster
			Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE			1 (bases 1 to 839)
AUTHORS			Genoscope.
TITLE			Direct Submission
JOURNAL			Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)
COMMENT			Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pbeloBAC11.
FEATURES			Location/Qualifiers
source		1..839	/organism="Drosophila melanogaster"
			/plasmid="pBelOBAC11"
			/db_xref="taxon:7227"
			/clone_lib="DrosBAC"
			/clones="BACN07E20"
			/note="end : Sp6"
BASE COUNT	511 a	54 c	107 g
ORIGIN		85 t	82 others
Query Match		6.3%;	Score 72.4; DB 229; Length 839;
Best Local Similarity		41.9%;	Pred. No. 1.le-06;
Matches	226;	Conservative	38; Mismatches 274; Indels 2; Gaps 1;
Qy	40	gaaaaagtacagaacgcgtgctgacaagtttgatgatagaaaacaatacgtttaaacacttaag	99
Db	76	GTATAAAWACMTGAAGTAGAATATATWTTMGCAAGGAAAAAAAAMMMTAAAGCACAWAAA	135
Qy	100	aatagtgacttaagtttcaataaataaagcgcttaaatgattcataatgatgagttaactgaa	159
Db	136	AAAAAAAAAAWWAAAAAAAAAGAAAAARAARAAAAAARAGAAAAAAMMAAAAAA	195
Qy	160	gagttgagtaatgctaagagaaaactacgtggatccgcgctgactaggggtacaataaat	219
Db	196	AAAGAACAAAAAARAARGAGAGMAAAAAAARAARAAAAAAMMAAAAAA	255
Qy	220	gaccgcgaagagcaagaagctcttgacaagtatgactagaacaacctgacttaaaa	279
Db	256	RAAAAAAAMMCGGAARGGGAAAAAAMMAAAAGAACMACMAAAAGAAAAARAAGAAAAA	315
Qy	280	actaagaatg--aaggggttaaactgagaatgaaggggttaaactgagaatgaaggt	337
Db	316	AAAAAAAAAAMAAGCAAFARAAAAARAARAAACCAAAAAAARAAGGAAGAAGCA	375
Qy	338	taaaactgagaatgaaggggttaaactgaggtgcacagagtggttccttaggggacgg	397
Db	376	AAAAAAGGAAAAAMWAAGCAGAAAAAARGAACRGGAARAAAAAARAAGAACGAAAA	435

QY 437 atgacgtagagaactctatgtttacaaagctaatatgacaaagttaccatggagagtgctt 496
 Db 1231 TCAAGGTATATAAAATAATATAATAAAATAATATAATAAAATAATATAATAAAATA 1172
 QY 497 atactaggcatccgcagagaataagcttaaaaaattattacgactcttgacgcacaaag 556
 Db 1171 AAATTATATAATAATATAATGAGAAAAAAGAGAAAAAAGAAAAAATAATAATAAAA 1112
 QY 557 aactgaattacacacacagatgagaagttatctctgcgaacgctggtgtaactcta 616
 Db 1111 AAAGGAAAGAAAGAGAGTAGATAATTTTAAANNAANNAAGATNAATAAATAAAT 1052
 QY 617 ggaagttatagaagattctgcagacaaactccgcacatacaaaataacgtttacgtc 676
 Db 1051 AATTAATAAGATTATTAATAATAAAATAATTAACAANAATTAATTAATAATAAAAA 992
 QY 677 acgaaacacagagcttaaaacgagatttagagaatgcaatggaagttgcaggaagatt 736
 Db 991 AAAAAAATAAAAGAAAAAAGAAAAAATAAAATAATAATAATAATAATAATAATAA 932
 QY 737 ttaagagagctggtaccttctgttagatcaggtttacacaaattatatatacaataatagta 796
 Db 931 TTAAGGNAATAAATTATNNNTAANNNTGGAATAATAAAAGATAAAAAAAGAG 872
 QY 797 attacacaaataatgcacaaagctgcagacttgacctgagacacaaaggctgaatc 856
 Db 871 AAAAAAATAAAAGAAAAAATAAAANNAANNAANNNNNNNNNAANNAANNAANNA 812
 QY 857 taaagccctaatgattggtgagaggtctgttacaagagtttaaatcagatgcgca 916
 Db 811 AANANNNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 752
 QY 917 ctggtctcagacagatactctggaagagtcacaaagctgctgacaaagtttgagatag 976
 Db 751 NNANNNNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 692
 QY 977 aacacaaactgtaaaacttaagatactgacttaagtttaataataaagcgttaaaag 1036
 Db 691 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 632
 QY 1037 atcataatgtagttaaactgagaggtgagtaatgctaaagagaaactgtaaaatg 1096
 Db 631 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 572
 QY 1097 ataaatcactatctgaaagcgtgtaaaattcaagagatttagaggcagcgttaagtaaa 1154
 Db 571 ANAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 514

RESULT 4
 LOCUS BE420745 1885 bp mRNA EST 24-JUL-2000
 DEFINITION HW002.802 ITSC HMM Barley Leaf Library Hordeum vulgare cDNA clone
 ACCESSION BE420745
 VERSION BE420745.1 GI:9418588
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
 Hordeum
 1 (bases 1 to 1885)
 REFERENCE
 AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
 S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
 Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
 Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
 Sorrells,M., Warburton,M. and Wenzel,G.
 TITLE International Triticeae EST Cooperative (ITREC): Production of
 Expressed Sequence Tags for Species of the Triticeae
 JOURNAL Unpublished (2000)

COMMENT Contact: Herrmann RG
 Botanisches Institut der LMU
 Menzinger Str. 67, D-80638 Munchen GERMANY
 Fax: 49 30 171683
 Email: herrmann@botanik.biologie.uni-muenchen.de
 International Triticeae EST Cooperative (ITREC)
 http://wheat.pw.usda.gov/genome.
 Location/Qualifiers
 FEATURES
 source
 1. 1885
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone="HMM002.802"
 /clone_lib="ITREC HMM Barley Leaf Library"
 /tissue_type="leaf"
 /dev_stage="14 day old"
 /note="vector: pBluescriptSK(-); 850 bp average insert size."
 BASE COUNT 1138 a 219 c 212 g 176 t 140 others
 ORIGIN
 Query Match 6.1%; Score 70.6; DB 166; Length 1885;
 Best Local Similarity 37.5%; Pred. No. 3.2e-06;
 Matches 370; Conservative 0; Mismatches 617; Indels 0; Gaps 0;
 QY 153 aactgaaggttgatgtaatgctaaagagaactacgtgacgcgtgactaggggtac 212
 Db 890 AAAAAAATAAAAGAAAAAATAAAANNAANNNNNNNNNNNNNNNNNNNNNNNNNNN 949
 QY 213 ataaatgaccgcgcaagagcaaaagagctcttgacaagtatgagctagaaacacatga 272
 Db 950 NNN 1009
 QY 273 cttaaaactgaatgaaggtttaaaactgagaaggttaaaactgagaaggttaaaactgaga 332
 Db 1010 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1069
 QY 333 aggggttaaaactgagaaggtttaaaactgagaaggttcgacagagtgcttcctagggg 392
 Db 1070 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1129
 QY 393 gacggttagaaacccgcgcaagcagcagagactcttcaacagtagcagtagagaactc 452
 Db 1130 NNAAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1189
 QY 453 tatgttacaagctaatgacaaagttaccatggagtgcttatactactacgcacgcc 512
 Db 1190 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1249
 QY 513 agaagataagcttaaaaaattattgacgatcttgacgcgcaaaagacatgattacaaca 572
 Db 1250 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1309
 QY 573 acagaatgagaagttatctctgcgagaacggtgtagttaaactcctagggaggttagaaga 632
 Db 1310 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1369
 QY 633 tcttcgacgcaaacatccgcgaataataatactattacgttcacgaaacagagactt 692
 Db 1370 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1429
 QY 693 aaaaagcagattagagaatgcaatggaagttgcaggaagagattttaagagagctgggtac 752
 Db 1430 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1489
 QY 753 ctgttagatcaggtttacacattatatacaataataataatagtaataaccacacataa 812
 Db 1490 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1549
 QY 813 tgcacaagctggcagactgacccagacaaagcgtgaatctctaaagccttaagta 872
 Db 1550 AAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAA 1609

QY	873	ttgggctagagcgtgttacaaaggttaaatatogtgcgactaggtctcagacaga	932
Db	1610	AA	1669
QY	933	tactctggaaaggtacaaagcgtgtgcacagtttgagatagaacacaaatcggttaa	992
Db	1670	AA	1729
QY	993	acttaagaatagtgacttaagttttaataataaagcgtttaaagatcatatgatgatt	1052
Db	1730	AA	1789
QY	1053	aactgaagaggttgagtaagtctcaagagaaactacgttaaaatgataaatacactatcga	1112
Db	1790	AA	1849
QY	1113	aaaagctagtaaaattcaagaattaga	1139
Db	1850	AA	1876
RESULT	5		
LOCUS	BF264918	1282 bp mRNA	17-NOV-2000
DEFINITION	HV_CEA0010L23f Hordeum vulgare seedling green leaf EST library		
ACCESSION	HV_CEA0010L23f	[Erysiphe infected & control] Hordeum vulgare cDNA clone	
VERSION	BF264918		
KEYWORDS	EST		
SOURCE	barley		
ORGANISM	Hordeum vulgare		
REFERENCE	1 (bases 1 to 1282)		
AUTHORS	Wing, R., Close, T.J., Klein, H., Wise, R., Begum, D., Frisch, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kern, D., Palmer, M., Rambo, T., Saski, C., Schwartz, J., Simmons, J., Choi, D.W., Main, D. and Wood, T.		
TITLE	Development of a genetically and physically anchored EST resource		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Wing RA Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: AATTAACCTCCTACTAAGGG High quality sequence start: 354 High quality sequence stop: 806		
FEATURES	Location/Qualifiers		
source	1. 1282		
	/organism="Hordeum vulgare"		
	/cultivar="C16155 (M1a13)"		
	/db_xref="taxon:4513"		
	/clone="HV_CEA0010L23f"		
	/clone_lib="Hordeum vulgare seedling green leaf EST library HV_CEA0004 (Erysiphe infected & control)"		
	/tissue_type="seedling green leaf"		
	/lab_host="TJC121"		
	/note="vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	6 a 0 c 0 g 524 t 752 others		
ORIGIN			
Query Match	6.0%	Score 69	DB 146
Best Local Similarity	27.5%	Pred. No. 7e-06	Length 1282
Mismatches	0	Mismatches 506	Indels 0
Gaps	0	Gaps	0

AUTHORS Roest-Crollius, H., Jaillon, O., Desilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W., and Weissbach, J.

TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1036)

AUTHORS Genoscope.

JOURNAL Direct Submission

COMMENT Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES Location/Qualifiers

Source 1..1036

 /organism="Tetraodon nigroviridis"

 /db_xref="taxon:99883"

 /clone="036N11"

 /clone_lib="C"

 /note="Genoscope sequence ID : C0BG036CG06SP1-end ; PUC-ori"

BASE COUNT 1005 a 6 c 7 g 17 t 1 others

ORIGIN

Query Match 5.9% Score 68; DB 231; Length 1036;
 Best Local Similarity 41.9% Pred. No. 1.le-05;
 Matches 407; Conservative 0; Mismatches 565; Indels 0; Gaps 0;

QY 183 actacgtggatccgcgtgactgaggggtacataaataatgaccgcgaagcaagaagc 242
 Db 9 ATTTGGTTATATTCCTTGACTTATTAAACCAAAAAAAAAAAAAAAAAAAAAA 68

QY 243 tcttgacaagtatgagctagaacacatgactctaaacctaagaatgaagggttaa 302
 Db 69 AA 128

QY 303 tgagaatgaagggttataaaactgagatgaagggttaaaactgagaatgaagggttaa 362
 Db 129 AA 188

QY 363 aactgaggtcgacagagtgttctctagggcggtagaaacccgcgaacgacgaga 422
 Db 189 AA 248

QY 423 actcttaacaagtatgacgtagagaactctatgtttacagctaataatgacagttacc 482
 Db 249 AA 308

QY 483 atggagagtcgttatactaggcctacgcgaagataagcttaaaaaaatttgacga 542
 Db 309 AA 368

QY 543 tcttgacgcacaagaacatgaattacaacaacagaatgagaagttatctctgcgaac 602
 Db 369 AA 428

QY 603 tgatggaatcctagggaagtattagaagatcttgcagcaaacatccgcgaatacaaaa 662
 Db 429 AA 488

QY 663 tatatgcttacgtcacgaacaaaggacttaaaagcgagattagagaatgcaatggaagt 722
 Db 489 AA 548

QY 723 tgcaggaagagattttaagagagctggttacctgttagatcaggttacacaattatatac 782
 Db 549 AA 608

QY 783 taacataatagtaattaccacaataatgacaaagctggcagacttgacctgagaca 842
 Db 609 AA 568

QY 843 aaaggctgaatctataaaggccttaagtattggctgagaggctgttataagagttaaa 902
 Db 669 AA 728

QY 903 tatcgtgtcgcgacttagtctcagacagatactctggaagataagcaacgtgctga 962
 Db 729 AA 788

QY 963 caagtttgatgagaaacaatacgtttaaaacttaagaatagtgacttaagtttaataa 1022
 Db 789 AA 848

QY 1023 taaagcgttaaaagatcataatgatgatgaactgaagagtgagtaattcctaagagaa 1082
 Db 849 AA 908

QY 1083 actacgttaaaatgataaatacactatctgaaagagctagtaaaattcaagaattagagc 1142
 Db 909 AA 968

QY 1143 acgtaagtataaa 1154
 Db 969 AAAAAAAAAAAAAA 980

RESULT 7

BF264948 1106 bp mRNA EST 17-NOV-2000

LOCUS HV.CEa0010N07f Hordeum vulgare seedling green leaf EST library

DEFINITION HV.CEa0010N07f (Erysiphe infected & control) Hordeum vulgare cDNA clone

ACCESSION BF264948

VERSION BF264948.1 GI:11195942

KEYWORDS EST.

SOURCE barley.

ORGANISM Hordeum vulgare

REFERENCE 1 (bases 1 to 1106)

AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Yu, Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo, T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and Wood, T.

TITLE Development of a genetically and physically anchored EST resource for barley genomics

JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7268
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: AATTAACCTCACTAAAGGG
 High quality sequence start: 190
 High quality sequence stop: 895.

FEATURES Location/Qualifiers

source 1..1106

 /organism="Hordeum vulgare"

 /cultivar="C116155 (M1a13)"

 /db_xref="taxon:4513"

 /clone="HV.CEa0010N07f"

 /clone_lib="Hordeum vulgare seedling green leaf EST library HV.CEa0010N07f"

 /tissue_type="seedling green leaf"

 /lab_host="TJC121"

 /note="Vector: lambdaZAP; Site_1: EcorI; Site_2: XhoI"

BASE COUNT 817 a 57 c 46 g 104 t 82 Others

ORIGIN

Tue May 15 07:27:16 2001

Brassicales: Brassicaceae; Arabidopsis.									
1 (bases 1 to 1223)									
Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and									
Ecker,J.									
BAC End Sequences at ATGC									
Unpublished (1997)									
Contact: Ecker J.									
Arabidopsis Thaliana Genome Center									
University of Pennsylvania									
Dept. of Biology, University of Pennsylvania, Philadelphia, PA									
19104									
Tel: 215-898-9384									
Fax: 215-898-8780									
Email: jecker@atgenome.bio.upenn.edu									
Seq primer: Sp6									
Class: BAC ends									
High quality sequence start: 199									
High quality sequence stop: 271.									
Location/Qualifiers									
1..1223									
/organism="Arabidopsis thaliana"									
/strain="Columbia"									
/db_xref="taxon:3702"									
/clone="T24D11"									
/clone.lib="TAMU"									
/sex="hermaphrodite"									
/note="vector: BelOBACII; Site_1: HindIII; Site_2: HindIII									
; Produced by Rod Wing"									
BASE COUNT 70 a 50 c 53 g 678 t 372 others									
ORIGIN									
Query Match 5.7%; Score 66.4; DB 227; Length 1223;									
Best Local Similarity 28.4%; Pred. No. 2.8e-05;									
Matches 283; Conservative 0; Mismatches 713; Indels 0; Gaps 0;									
Qy	71	agatagaaacatacgtttaaaccttaagataagtagcttaagtttttaataaagaagcgt	130						
Db	1216	AA	1157						
Qy	131	taaaagatcataatgatgagtttaactgaagagtgtagtgaagctgaagagaaactcagtg	190						
Db	1156	AA	1097						
Qy	191	gatccgcgtgactagggttacaataaatgacccgcgaagagcaaaagagctcttgaca	250						
Db	1096	AA	1037						
Qy	251	agtgatgagctagaaac	310						
Db	1036	AA	977						
Qy	311	aagggttaaaac	370						
Db	976	AA	917						
Qy	371	tcgacagagtgcttcctaggggagcggttagaaccgagcagcagcagcagcagcagc	430						
Db	916	NN	857						
Qy	431	acaagtacgctagagaaactctctgttacaagcttaataatgacaagttaccatgagag	490						
Db	856	AA	797						
Qy	491	tgcgttactaggcatcgcgcgaagataagcctaaataattattgacgctcttgacg	550						
Db	796	NN	737						
Qy	551	caaaagacatgaattacaacacagagtgagagttatctctgcagacggtgatggtta	610						
Db	736	AA	677						
Qy	611	atcctagggaagttagaagatctcttgacgcaaacacacacacacacacacacacac	670						

Db	789	DRDAAATTTTWTITTTTDTDDKWKTTDTWTAAADRTWDRDDDDDRDRAGTAGRKWRRT	730
Qy	959	ctgacaagtttgagatagaaacaatacgttaaaacttaagaactgaagcttaagtttta	1018
Db	729	WKRWRKRDRTWDDADADDTARDDRRGGDDGADAGKKGKTKRRRRDRATWDRTDANW	670
Qy	1019	ataataaagcgtttaaagatcataatgagcttaactgaagaggttgagtaagtctaag	1078
Db	669	ADAAWTTTDTDDMDKDRRRKRGARRRRRTTAAAMDWMTWAKANDWAKWKTTRADRW	610
Qy	1079	agaactacgttaaaatgataatcactctctgaaaagaagct	1119
Db	609	DRWAATWTDARKARDWAKARAWRARRDRARAARADREWT	569
RESULT 11			
LOCUS	CNS017RP	1101 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37J10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL108415	GI:5628719	
VERSION	GSS.		
KEYWORDS	fruit fly.		
SOURCE	Plasmid Drosophila melanogaster		
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Direct Submission		
TITLE	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :		
JOURNAL	Bp 191 91006 EVRY cedex - FRANCE. (E-mail : seqref@genoscope.cns.fr		
COMMENT	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Pavan. It has been constructed in the vector pbelOBAC11.		
FEATURES	Location/Qualifiers		
source	1..1101		
	/organism="Drosophila melanogaster"		
	/plasmid="pbelOBAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN37J10"		
	/note="end : SP6"		
BASE COUNT	258 a 174 c 277 g 120 t	272 others	
ORIGIN			
Query Match	5.6%	Score 65;	DB 229; Length 1101;
Best Local Similarity	27.8%;	Pred. NO. 5.7e-05;	
Matches	107;	Conservative 104;	Mismatches 174; Indels 0; Gaps 0;
Qy	39	ggaagaagtacgaacgctgcgaagtttgagatagatagaacaacatcagtttaaaacttaa	98
Db	660	SGRAARASGRGAGAAARAARAGAGAGARRARARARARARARARARARAGAGA	719
Qy	99	gaatagtgacttaagtttaataataaagcgttaaaagatcataagtgagcttaactga	158
Db	720	RRRGARGAGAGAGAAARARARARARAGSGGRRARRARARARARARARARARARAR	779
Qy	159	agatgtgagtaagtctaaagagaaaactcgtggatccgcgctgactaggggtacaataa	218
Db	780	AGAAAGGAGAAARRRRRAGRRGGRRGRRGRRGGGARGAGAGAGRGGARAAAARRAR	839
Qy	219	tgaccgcgaaagacaaagaagctcttgacaagtatgagctagaacaaacatgacttaa	278

RESULT 10
 CNS0039G 1101 bp DNA GSS 03-JUN-1999
 Drosophila melanogaster genome survey sequence TEF3 end of BAC #
 LOCUS BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 DEFINITION
 AL063921
 AL063921.1 GI:4941778
 GSS:
 fruit fly.
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE
 1 (bases 1 to 1101)
 Genoscope.
 AUTHORS
 Direct Submission
 TITLE
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT
 - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazuhiro Osoegawa and
 Aaron Mammoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 p1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 found at Location/Qualifiers
 1. 1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR08K10"
 /note="end : TEF3"
 BASE COUNT 201 a 64 c 131 g 202 t 503 others
 ORIGIN
 FEATURES
 source
 1. 1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR08K10"
 /note="end : TEF3"
 BASE COUNT 201 a 64 c 131 g 202 t 503 others
 ORIGIN
 Query Match 5.6% Score 65.4; DB 229; Length 1101;
 Best Local Similarity 15.98; Pred. No. 4.6e-05;
 Matches 83; Conservative 249; Mismatches 187; Indels 2; Gaps 1
 Qy 601 ggtgatgtaactctagggaagttatagagagattcgcagcaaacatccgcgaataca 660
 Db 1089 RDRTRKDDWTKWMTWKDRADRRWAGDADRWADDDGAGTGWTTATWWWWWATWDTW 1030
 Qy 661 aatatacattcagtcacgaaaaaacagacttaaacgcagattagagaatgcgaatg 720
 Db 1029 WDKWWWATAAKTDTATWTTTAWRADWAGRDGACGKRDRDAATDADGAGRRDGRKKD 970
 Qy 721 gttgcagaagagatttaagagagctggtagcttctgttagatcaggttta--cacaaatt 778
 Db 969 KKDRKDDGDDDKGGKKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 910
 Qy 779 atactaaacataatagtaattaccacaataataatgcacaaagctggcagagctgacctga 838
 Db 909 DGGKGDADDTDGTKDDDDKDDWDDKAKGTWGDATWAAATATWWWGWADADWMTWDA 850
 Qy 839 gacaaagcgtgaatactctaaaagccttaagtgttggtgctgagagcgtgttacaagagt 898
 Db 849 AADDWWADRRWADAWKDDAWAWGARTADRRDWDGDRACKRGARKRRDRKRRADKRDAA 790
 Qy 899 taaatatcgatgtcgcagactgtctccagacagatactctcgaaaagatcagaagaactg 958


```

Query Match          5.6%   Score 64.8;   DB 223;   Length 853;
Best Local Similarity 78.0%   pred. No. 6.1e-05;
Matches 78;   Conservative 0;   Mismatches 22;   Indels 0;   Gaps 0;
QY  269  atgaactttaaaactaagaatgaagggttaaaaaactgagaatgaagggttaaaaaacttgaga 328
      ||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 14

CNS00599/c	LOCUS	DEFINITION	CNS00599	1036 bp	DNA	GSS	03-JUN-1999
		Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR1116 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.					
ACCESSION	AL057797	GI:4932579					
VERSION	AL057797.1						
KEYWORDS	GSS.						
SOURCE	fruit fly.						
ORGANISM	Drosophila melanogaster						
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.						
REFERENCE	1 (bases 1 to 1036)						
AUTHORS	Direct Submission						
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage :						
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr						
COMMENT	- web : www.genoscope.cns.fr) Determination of this BAC end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain v2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.						
FEATURES	Location/Qualifiers						
SOURCE	1..1036						
	/organism="Drosophila melanogaster"						
	/db_xref="taxon:722"						
	/clone_lib="RPCI-98"						
	/clone="BACR1116"						
	/note="end : TET3"						
BASE COUNT	64 a 56 c 41 g 701 t 174 others						
ORIGIN							
	Query Match						
	Best Local Similarity 37.2%; Score 64.6; DB 229; Length 1036;						
	Matches 262; Conservative 78; Mismatches 356; Indels 9; Gaps 2;						
QY	459 acaagctaatatgacaagttaccatgagagtgcggttatactataggcatatcagcagaaga 518						
Db	: :						
	994 ACAAGACRCARMACGHCWCYAACGCAGWTGTGARTATAAAATKYGAAWCKTRAWAGAK 935						
QY	519 taagtctaaaaaattattgcagctcttgacgcacaaagacatgatctacacacacagaa 578						
Db	: :						
	934 TWAASAAARASTCWAAMGYCCASATWAAASAAATWARGWRAAASCYGAGMAHAT 875						
QY	579 tgagaagttatctctgcagaacggttgagtgtaactcctagggaagttatagaagattctgc 638						
Db	: :						
	874 SSAAAGCGWSACGWKASACAYGYAAATAAAAAAAGAAKAGTCTCTGYTCYKCYGT 815						
QY	639 agcaaaccaatcccgcaatacacaataatacgttttaagtcacgaaaaacaagact-taaaag 697						
Db	: :						
	814 CTCAGKAKACGCGYGAAAAAAMATTCAGHAACAYAAAAAAAAYGASTCTAAAT 755						
QY	698 cgaagtagaagatgcaatgggaagttgcggaagagattttaagagagctgggtacctgt 757						
Db	: :						
	754 AGRAAAATATYCCSMAAAHAKACTAGRAAAAAAARAGAAVAAWAAAAAGVHBAKTKK 695						
QY	758 tagatcaggttacacaattatatactataacataatagtagtaattaccacaa-----ta 809						
Db	: :						
	694 CKKAAAMAVYTCNTAATAABKAKCTTWAAMAAAAAAMVMAAMRCACAAAMAAAAAYGISTS 635						

Search completed: May 13, 2001, 04:29:58
Job time: 9128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:07 ; Search time 21.02 seconds
(without alignments)
1041.557 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 1954

Sequence: 1 ACWVATRSQDTFLKRVQRA.....KNDKSLSEKSKIQELEAR 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID56/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1954	100.0	383	20 Y04368	Streptococcal hexa
2	966	49.4	247	21 B03114	S. pyogenes hybrid
3	961	49.2	247	15 R50993	Recombinant M24-M5
4	932.5	47.7	274	15 R50994	Recombinant M24-M5
5	932.5	47.7	274	21 B03115	S. pyogenes hybrid
6	622	31.8	247	15 R50997	Recombinant M19-M6
7	622	31.8	247	21 B03119	S. pyogenes hybrid
8	622	31.8	343	15 R50999	Recombinant tetra
9	622	31.8	343	21 B03121	S. pyogenes hybrid
10	576	29.5	173	15 R50998	Recombinant M24-M5
11	576	29.5	173	21 B03120	S. pyogenes hybrid

12	392	20.1	287	15 R50992	Recombinant M24-M5
13	392	20.1	287	21 B03113	S. pyogenes hybrid
14	338	17.3	135	15 R51001	M24-M5-M6-M19-M3-M
15	337.5	17.3	135	21 B03124	S. pyogenes octava
16	335	17.1	305	15 R50996	Recombinant M24-M5
17	335	17.1	305	21 B03117	S. pyogenes hybrid
18	321	16.4	187	15 R50995	Recombinant M24-M5
19	321	16.4	187	21 B03116	S. pyogenes hybrid
20	267	13.7	67	15 R51000	M24-M5-M6-M19 mult
21	267	13.7	67	15 R51000	S. pyogenes hybrid
22	257.5	13.2	483	18 W08927	Type-6 M-protein.
23	256	13.1	441	14 R41780	Streptococcus pyog
24	249	12.7	441	10 P90955	M6 streptococcal p
25	220.5	11.3	327	18 W20901	H. pylori secreted
26	207.5	10.6	980	21 B18294	Plasmodium falcipa
27	205	10.5	1392	20 Y06999	Restin protein seq
28	203	10.4	42	16 R74258	SSAL streptococcal
29	202	10.3	1427	12 R10534	Human 160kD mediat
30	198.5	10.2	2954	20 Y01632	Amino acid sequenc
31	196	10.0	1979	21 B18171	Plasmodium falcipa
32	193.5	9.9	561	19 W63043	Streptococcus uber
33	190.5	9.7	753	21 B08316	A human M-phase p
34	189.5	9.7	100	6 P50295	Sequence encoded b
35	189	9.7	1886	19 W54241	Rattus norvegicus
36	188.5	9.6	2482	16 R72826	Human mitosis. Ho
37	188.5	9.6	2482	19 Y23996	Human mitosis. Ho
38	188	9.6	472	21 Y23943	Human cancer assoc
39	187.5	9.6	979	21 G40045	Arabidopsis thalia
40	187.5	9.6	1019	21 G40044	Arabidopsis thalia
41	187.5	9.6	1058	21 G40044	Arabidopsis thalia
42	187.5	9.6	1342	21 G31251	Arabidopsis thalia
43	187.5	9.6	1382	21 G31250	Arabidopsis thalia
44	187.5	9.6	1421	21 G31249	Arabidopsis thalia
45	187.5	9.6	3248	17 R99795	Kinetochore protei

ALIGNMENTS

RESULT 1
Y04368
ID Y04368 standard; Protein; 383 AA.
XX
AC Y04368;
XX
DT 23-JUN-1999 (first entry)
XX
DE Streptococcal hexavalent M protein vaccine.
XX
KW Streptococcal; hexavalent emm gene; multivalent; vaccine; Group A;
KW immunogenic; immune response; pathogen; fusion protein.
XX
OS Streptococcus sp.
XX
PN WO9913084-A1.
XX
PD 18-MAR-1999.
XX
PF 14-SEP-1998; 98WO-US19100.
XX
PR 12-SEP-1997; 97US-0058635.
XX
(IDVA-) ID VACCINE.
XX
Dale JB;
WPI; 1999-215066/18.
N-PSDB; X33103.
Immunogenic fusion protein derived from group A streptococci
Example 1; Fig 7; 50pp; English.

CC The present invention describes an immunogenic fusion polypeptide (I)
 CC that stimulates an immune response against group A streptococci (GAS).
 CC (I) comprises: (a) at least 2 peptides (Ia) from GAS, at least 10 amino
 CC acids long and able to induce a specific immune response; and (b) a
 CC peptide (Ib), C-terminal to (Ia) that protects the immunogenicity of the
 CC component described in (a) but is not essential for stimulation of the
 CC immune response. Vaccines containing (I) are used to protect against GAS
 CC infections, specifically those caused by Streptococcus pyogenes, e.g.
 CC pharyngitis, pyoderma, toxic shock syndrome, deep tissue invasion, sepsis
 CC and acute rheumatic fever. (I) have improved immunogenicity and do not
 CC generate antibodies that are cross-reactive with human tissues. The
 CC present sequence represents a hexavalent M protein vaccine.

XX Sequence 383 AA;

Query Match 100.0%; Score 1954; DB 20; Length 383;
 Best Local Similarity 100.0%; Pred. No. 1.2e-124;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACMVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSEFNKALKDHNDLTELSELSNAKE 60
 DB 1 acmvatrsqdtlekvgeradkfeientlklknsdlsefnkalkdhndlteelsnake 60
 QY 61 KLKRSVTRGTINDPQRAKEALDKYELENHDLKTKNEGLKTEENEGKLTENEGL 120
 DB 61 klkrsavtrgtindpqrakealdkyelenhdltkneglkteneglkteneglkteneg 120
 QY 121 KTEVDVFPRTGVENPKARLLNKYDVENSMLQANDKLPWRVRYTRHPEDKLLKIID 180
 DB 121 ktevdvfpgrtgvnpkarellnkkydvensmlqandklpwrvrytrhtpedklkiiid 180
 QY 181 DLDAKEHELOQNEKLSLQNGDGNPREVIEDLAANNPAIGNIRLRHNKDLKARLENAME 240
 DB 181 dlakeheiqngneklslqngdgnprevedlaannpaignirlnrhenkdikarlename 240
 QY 241 VAGRDFKRAGLTLDQVTLQYTKHNSNTQQYNAQAGRLDLRQKAEYKGLNDWAEPLQL 300
 DB 241 vagrdfkragtldqvtlqytkhnsntqqynaqagrldlrqkaeykglndwaerllqel 300
 QY 301 NIDVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSEFNKALKDHNDLTELSELSNAKE 360
 DB 301 nidvatrsqdtlekvgeradkfeientlklknsdlsefnkalkdhndlteelsnake 360
 QY 361 KLKNDKSLSEKSKIOLEARK 383
 DB 361 klrkdkslsekskiobleark 383

RESULT 2

B03114
 ID B03114 standard; Protein; 247 AA.
 XX
 AC B03114;
 XX
 DT 10-OCT-2000 (first entry)
 XX
 DE S. pyogenes hybrid M protein (M24-M5-M6-M19), SEQ ID NO:4.
 XX
 KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia.
 XX
 OS Streptococcus pyogenes.
 OS Synthetic.
 XX
 PN US6063386-A.
 XX
 PD 16-MAY-2000.
 XX
 PF 15-SEP-1997; 97US-0937271.

XX 16-SEP-1992; 92US-0945954.
 PR (UYTE-) UNIV TENNESSEE RES CORP.
 PA
 PI Lederer JW, Dale JB;
 XX WPI; 2000-364475/31.
 DR N-PSDB; A57894.
 XX
 PT New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever
 XX
 PS Disclosure; Fig 4A-B; 62pp; English.
 XX
 CC The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci. Type A streptococci
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic
 CC heart disease. In a patient with ARF, antibodies formed during a group A
 CC streptococcal infection are also cross-reactive with heart tissue which
 CC indicates that the streptococci and host tissue contain similar antigenic
 CC motifs. The new multivalent vaccines are capable of raising sero-specific
 CC antibodies against various serotypes of group A streptococci which are
 CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
 CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
 CC Streptococcus pyogenes M proteins generated in the disclosure of the
 CC invention.
 XX Sequence 247 AA;

Query Match 49.4%; Score 966; DB 21; Length 247;
 Best Local Similarity 78.9%; Pred. No. 3.9e-58;
 Matches 195; Conservative 0; Mismatches 0; Indels 52; Gaps 1;

QY 3 MVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSEFNKALKDHNDLTELSELSNAKE 62
 DB 1 mvatrsqdtlekvgeradkfeientlklknsdlsefnkalkdhndlteelsnake 60
 QY 63 R-----GSAVTRG 70
 DB 61 rkndkslseksknglearkadkalegannfstadsakiktlaekadlegsavtrg 120
 QY 71 TINDPQRAKEALDKYELENHDLKTKNEGLKTEENEGKLTENEGLKTEVDVFP 130
 DB 121 tindpqrakealdkyelenhdltkneglkteneglkteneglkteneglktevdvfp 180
 QY 131 GTVENPDKARELINKYDVENSMLQANDKLPWRVRYTRHTPEDKLLKIIDDLDAKEHELO 190
 DB 181 gtvenpdkarellnkkydvensmlqandklpwrvrytrhtpedklkiiiddlakahehlo 240
 QY 191 QQNEKLS 197
 DB 241 qgneklis 247

RESULT 3
 R50993

Query Match 49.2%; Score 961; DB 15; Length 247;
Best Local Similarity 78.5%; Pred. No. 8.4e-58;
Matches 194; Conservative 0; Mismatches 1; Indels 52; Gaps 1;

PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes

PS Disclosure; Fig 6; 67pp; English.

XX The sequences given in R50992-1001 represent hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC pKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
CC to an immunogenic carrier.

XX Sequence 274 AA;

Query Match 47.7%; Score 932.5; DB 15; Length 274;
Best Local Similarity 71.2%; Pred. No. 8e-56;
Matches 195; Conservative 0; Mismatches 0; Indels 79; Gaps 3;

Qy 3 MVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALKDHNDLTELSELSNAKEKL 62
Db 1 mvatrsqdtlekvqeradkfeienntklknsdlsfnnkalkdhndelteelsnakekl 60
Qy 63 R-----
Db 61 rkndkslsekskngelarkadlekalegamnfstadsakiktaleakadlegspgnpa 120
Qy 64 --GSAVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTENEGLKTENEGLK 121
Db 121 vpgsavtrgtindpqrakealdkyelenhdltkneglkteneglkteneglk 180
Qy 122 TE-----VDRVPRGTVENPKARELLNKYDVNSMLQANNDKL-----PWR 163
Db 181 tevdpngnvpavpdrvfgtvenpdkarellnkdyvensmlqanndklpwpngnvpavpwr 240
Qy 164 VRYTRHTPEDKLKKIIDDLDLDAKEHELOQQNEKLS 197
Db 241 vrytrhtpedklkikiiddldakehehlgqnekl 274

RESULT 5

ID B03115
XX B03115 standard; Protein; 274 AA.

AC B03115;

XX 10-OCT-2000 (first entry)

DE S. pyogenes hybrid M protein (M24-M5-M6-M19), SEQ ID NO:6.

XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia.

XX Streptococcus pyogenes.

OS Synthetic.

XX US6063386-A.

XX

PD 16-MAY-2000.
XX 15-SEP-1997; 97US-0937271.
XX 16-SEP-1992; 92US-0945954.
XX (UYTE-) UNIV TENNESSEE RES CORP.
PA Lederer JW, Dale JB;
XX WPI; 2000-364475/31.
XX N-PSDB; A57895.
PT New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever.
XX Disclosure; Fig 6A-B; 62pp; English.

XX The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
CC Streptococcus pyogenes M proteins generated in the disclosure of the
CC invention.

XX Sequence 274 AA;

Query Match 47.7%; Score 932.5; DB 21; Length 274;
Best Local Similarity 71.2%; Pred. No. 8e-56;
Matches 195; Conservative 0; Mismatches 0; Indels 79; Gaps 3;

Qy 3 MVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSPNNKALKDHNDLTELSELSNAKEKL 62
Db 1 mvatrsqdtlekvqeradkfeienntklknsdlsfnnkalkdhndelteelsnakekl 60
Qy 63 R-----
Db 61 rkndkslsekskngelarkadlekalegamnfstadsakiktaleakadlegspgnpa 120
Qy 64 --GSAVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTENEGLKTENEGLK 121
Db 121 vpgsavtrgtindpqrakealdkyelenhdltkneglkteneglkteneglk 180
Qy 122 TE-----VDRVPRGTVENPKARELLNKYDVNSMLQANNDKL-----PWR 163
Db 181 tevdpngnvpavpdrvfgtvenpdkarellnkdyvensmlqanndklpwpngnvpavpwr 240
Qy 164 VRYTRHTPEDKLKKIIDDLDLDAKEHELOQQNEKLS 197
Db 241 vrytrhtpedklkikiiddldakehehlgqnekl 274


```

RESULT 6
R50997
ID R50997 standard; Protein; 247 AA.
XX
AC R50997;
XX
DT 02-NOV-1994 (first entry)
XX
DE Recombinant M19-M6-M4-M24 tetraivalent hybrid.
XX
KW Primer: PCR; amplify; polymerase chain reaction; construct; hybrid;
KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
KW restriction enzyme site; multivalent M protein; immunisation; group A;
KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
KW antibody; heart tissue; antigen; serotype; mucosal.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 2..38
FT /label= M19
FT Peptide 39..75
FT /label= M6
FT Peptide 76..135
FT /label= M5
FT Peptide 136..247
FT /label= M24
XX
PN W09406421-A.
XX
PD 31-MAR-1994.
XX
PF 15-SEP-1993; 93WO-US08703.
XX
PR 16-SEP-1992; 92US-0945954.
XX
PA (UYTE-) UNIV TENNESSEE RES CORP.
XX
PI Dale JB, Lederer JW;
XX
DR WPI; 1994-118122/14.
DR N-PSDB; Q45219.
XX
XX New immunogenic hybrid proteins derives from streptococcal M proteins
XX - induces opsonic antibodies, for protective immunisation against
XX against multiple group A streptococci serotypes
XX
PS Disclosure; Fig 9; 67pp; English.
XX
CC The sequences given in R50992-1001 represent hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC PKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
XX
SQ Sequence 247 AA;

```

```

Query Match 31.8%; Score 622; DB 15; Length 247;
Best Local Similarity 39.2%; Pred. No. 6.7e-35;
Matches 149; Conservative 10; Mismatches 17; Indels 204; Gaps 3;

QY 4 VATRSQTDITLEKVOBRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSENAKEKLR 63
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
40 vfrgtvnpdkarellnkdyvensmlqandnlv----- 74

QY 64 GSAVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTENEGTKTENEGTKTENEGKTE 123
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
75 -davgtrgtindpqrakealdkyelenhdltkneglktenegltktenegltkteneglkte 133

QY 124 VDRVFRGTGVNDPKARELLNKYDVENSMLQANDKLPWRVRYRTRHTPEDKLEKKIIDDL 183
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 -----pw----- 135

QY 184 AKEHELQQQNEKLSLQNGDGNPREVIEDLAANNPAIQIRLRHENKDLKARLENAMVAG 243
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
136 ----- 135

QY 244 RDKFRAGTLLDQVTLQYTKHNSYQQYNAQAGRLDROKAEYLGKLNDAERLLQELNID 303
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
136 ----- 135

QY 304 VATRSQTDITLEKVOBRADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSENAKEKLR 363
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
136 vatrsqtdtlekvqeradkfeienntlkksdisfnkalkdhndeltealsnakeklr 195

QY 364 KNDKSLSEKASKIOLEARK 383
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
196 knkdslekasknqleark 215

RESULT 7
B03119
ID B03119 standard; Protein; 247 AA.
XX
AC B03119;
XX
DT 10-OCT-2000 (first entry)
XX
DE S. pyogenes hybrid M protein (M19-M6-M5-M24), SEQ ID NO:13.
XX
KW Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia.
XX
OS Streptococcus pyogenes.
OS Synthetic.
PN US6063386-A.
XX
PD 16-MAY-2000.
XX
PF 15-SEP-1997; 97US-0937271.
XX
PR 16-SEP-1992; 92US-0945954.
XX
PA (UYTE-) UNIV TENNESSEE RES CORP.
XX
PI Lederer JW, Dale JB;
XX
DR WPI; 2000-364475/31.
XX
XX New immunogenic recombinant hybrid M protein comprising amino-terminal
XX peptide fragments of streptococcal M protein useful as vaccine against
XX rheumatic fever and infections leading to rheumatic fever
XX
XX Disclosure; Fig 9A-B; 62pp; English.
XX
XX The invention relates to multivalent immunogenic hybrid group A
XX
CC

```


	Peptide	2..115 /label= M24
FT FT	Peptide	116..173 /label= M5
XX XX	WO9406421-A.	
PN PD	31-MAR-1994.	
PP PF	15-SEP-1993; 93WO-US08703.	
PR PX	16-SEP-1992; 92US-O945954.	
PX PA	(UYTE-) UNIV TENNESSEE RES CORP.	
PI PI	Dale JB, Lederer JW;	
PD DR	WPI: 1994-I18122/14.	
DR DX	N-PSDB; Q45220.	
XX PT	New immunogenic hybrid proteins derives from streptococcal M proteins against multiple group A streptococci serotypes	
PT PP	- induces opsonic antibodies, for protective immunisation against	
PX PY	against multiple group A streptococci serotypes	
YY ZZ	Disclosure; Fig 10; 67pp; English.	
CC CC	The sequences given in R50992-1001 represent hybrid M proteins which contain the M24-M5-M6 and/or M19 subunits. These multivalent proteins were constructed using fragments of the 5' regions of emm genes that were amplified by PCR, ligated in tandem and expressed in pKK223.3. The amplified regions pref. encode protective and not tissue-cross-reactive epitopes, which can then be linked into one protein molecule. The recombinant hybrid protein may contain 113 N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and 35 from M19. Each section is linked by 2 amino acids specified by the respective restriction enzyme sites that were synthesised into the primers used to specify the PCR product. Multivalent M proteins such as this may be used for protective immunisation against Group A streptococci, which esp. cause rheumatic fever and rheumatic heart disease. Humoral antibodies raised against these proteins do not react with heart tissue antigens but are effective against many different serotypes. The multivalent proteins may also include sequences which induce mucosal antibodies and do not require coupling to an immunogenic carrier.	
CC CX	Sequence 173 AA;	
SQ SY	Query Match 29.5%; Score 576; DB 15; Length 173; Best Local Similarity 69.9%; Pred. No. 5.6e-32; Matches 121; Conservative 0; Mismatches 0; Indels 52; Gaps	
QY QZ	3 MVATRSOTDTLEKVOERADKFIEINNNTKLKNSDLSFNNAKDHDNDELTEELSNAKEKL 62 Db 1 mvatrsqdttlekvradkfieenntklknslsfnnkalhdndelteelssnakekl 60 	
QY QB	63 R-----GSAVTRG 70 	
Db DC	61 rkndksisekasngqelearkadlekalgamnfstadsakikltleaeakadlegsavtrg 120 	
QY QE	71 TINDPORAKALDYELNHDLTKNGGLKTENEGLKLTENEGLKTE 123 	
Db DE	121 tindpqraekaldyelenholcknkneglkteneeglkteneglkteneglikte 173 	
RESULT	11	
B03120	ID B03120 standard; Protein; 173 AA.	
XX AC	B03120:	
XX AD		
DX DT	10-OCT-2000 (first entry)	
XX EX		

QY 71 TINDPQRAKALDKYELNHDLTKNGLKTENGLKTENGLKTENGLKTE 123
 Db 121 tindpqrakealdkyelenhdltkkneglkteneglkteneglkteneglkte 173

RESULT 12
 R50992
 ID R50992 standard; Protein; 287 AA.
 XX AC
 XX R50992;
 XX
 DF 02-NOV-1994 (first entry)
 XX
 DE Recombinant M24-M5-M6.
 XX
 XX
 KW Primer; PCR; amplify; polymerase chain reaction; construct; hybrid;
 KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
 KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
 KW restriction enzyme site; multivalent M protein; immunisation; group A;
 KW streptococci; rheumatic fever; rheumatic heart disease; group A;
 KW antibody; heart tissue; antigen; serotype; mucosal.
 XX
 OS Synthetic.
 XX
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..262
 FT /label= M24
 FT Peptide 263..275
 FT /label= M5
 FT Peptide 276..287
 FT /label= M6
 XX
 PN WO9406421-A.
 XX
 PD 31-MAR-1994.
 XX
 XX 15-SEP-1993; 93WO-US08703.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Dale JB, Lederer JW;
 XX
 DR WPI; 1994-118122/14.
 DR N-PSDB; Q45214.
 XX
 XX New immunogenic hybrid proteins derives from streptococcal M proteins
 PT - induces opsonic antibodies, for protective immunisation against
 PT against multiple group A streptococci serotypes
 XX
 PS Disclosure; Fig 1; 67pp; English.
 XX
 CC The sequences given in R50992-1001 represent hybrid M proteins which
 CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
 CC proteins were constructed using fragments of the 5' regions of emm
 CC genes that were amplified by PCR, ligated in tandem and expressed in
 CC PKK223.3. The amplified regions pref. encode protective and not
 CC tissue-cross-reactive epitopes, which can then be linked into one
 CC protein molecule. The recombinant hybrid protein may contain 113
 CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
 CC 35 from M19. Each section is linked by 2 amino acids specified by
 CC the respective restriction enzyme sites that were synthesised into
 CC the primers used to specify the PCR product. Multivalent M proteins
 CC such as this may be used for protective immunisation against group A
 CC streptococci, which esp. cause rheumatic fever and rheumatic heart
 CC disease. Humoral antibodies raised against these proteins do not
 CC react with heart tissue antigens but are effective against many
 CC different serotypes. The multivalent proteins may also include
 CC sequences which induce mucosal antibodies and do not require coupling
 CC to an immunogenic carrier.

SQ Sequence 287 AA;
 Query Match 20.1%; Score 392; DB 15; Length 287;
 Best Local Similarity 100.0%; Pred. NO. 2.8e-19;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLFSFNKALKDHNDLTELSTNAKEKLR 363
 Db 2 vatrsqtdtlekvgeradkfeienntklknsdlfsfnkalkdhndelteelsnakeklr 61

QY 364 KNDKSLSEKASKIQLELARK 383
 Db 62 kndkslsekaskiqleleark 81

RESULT 13
 B03113
 ID B03113 standard; Protein; 287 AA.
 XX AC
 AC B03113;
 XX
 DF 10-OCT-2000 (first entry)
 XX
 DE S. pyogenes hybrid M protein (M24-M5-M6), SEQ ID NO:2.
 XX
 KW Multivalent hybrid M protein; group A streptococcus; serotype;
 KW immunogenic; sero-specific antibody; streptococcal infection;
 KW cross reactivity; vaccine; acute rheumatic fever; ARF;
 KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
 KW pneumonia.
 XX
 OS Streptococcus pyogenes.
 OS Synthetic.
 XX
 PN US6063386-A.
 XX
 PD 16-MAY-2000.
 XX
 XX 15-SEP-1997; 97US-0937271.
 XX
 PR 16-SEP-1992; 92US-0945954.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 XX
 PI Lederer JW, Dale JB;
 XX
 DR WPI; 2000-364475/31.
 DR N-PSDB; A57893.
 XX
 XX New immunogenic recombinant hybrid M protein comprising amino-terminal
 PT peptide fragments of streptococcal M protein useful as vaccine against
 PT rheumatic fever and infections leading to rheumatic fever
 XX
 PS Disclosure; Fig 1A-B; 62pp; English.
 XX
 CC The invention relates to multivalent immunogenic hybrid group A
 CC streptococcal M proteins comprising N-terminal peptide fragments of M
 CC proteins that elicit opsonic antibodies against multiple serotypes
 CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
 CC generated using the hybrid proteins are against one or more M protein
 CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
 CC The invention also encompasses a recombinant DNA molecule comprising a
 CC nucleotide sequence that encodes a multivalent hybrid M protein to
 CC method for immunising a mammal against streptococcal infections,
 CC comprising administering an immunogenic multivalent hybrid M protein to
 CC the mammal. The multivalent hybrid M proteins are useful for eliciting
 CC opsonic or protective antibodies to the M proteins of different serotypes
 CC of group A streptococci, and may therefore be used as vaccines to protect
 CC against and control infection by type A streptococci (strep throat),
 CC are not only responsible for streptococcal pharyngitis (strep throat),
 CC forms of pneumonia and a condition resembling toxic shock, but are also
 CC involved in the development of acute rheumatic fever (ARF) and rheumatic

CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
CC Streptococcus pyogenes M proteins generated in the disclosure of the
CC invention.

XX Sequence 287 AA;

Query Match 20.1%; Score 392; DB 21; Length 287;
Best Local Similarity 100.0%; Pred. No. 2.8e-19;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 VATRSQTDTLEKQVQRADKFEIENNTLKNSDLSFNKALKDHNDLTELSELSNAKEKL 363
DB 2 VATRSQTDTLEKQVQRADKFEIENNTLKNSDLSFNKALKDHNDLTELSELSNAKEKL 61

QY 364 KNDKSLSEKASKIOBLEARK 383
DB 62 kndkslsekaskioleark 81

RESULT 14
R51001
ID R51001 standard; Protein; 135 AA.
XX R51001;
XX
XX 02-NOV-1994 (first entry)
DE M24-M5-M6-M19-M3-M1-M18-M12 multivalent hybrid M protein.
XX
XX Primer: PCR; amplify; polymerase chain reaction; construct; hybrid;
KW M protein; M24; M5; M6; M19; subunit; tetraivalent protein; emm gene;
KW tandem; PKK223.3; protective epitope; tissue-cross-reactive epitope;
KW restriction enzyme site; multivalent M protein; immunisation; group A;
KW streptococci; rheumatic fever; rheumatic heart disease; humoral;
KW antibody; heart tissue; antigen; serotype; mucosal.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= M24
FT Peptide 19..35
FT /label= M5
FT Peptide 36..52
FT /label= M6
FT Peptide 53..69
FT /label= M19
FT Peptide 60..86
FT /label= M3
FT Peptide 87..103
FT /label= M1
FT Peptide 104..120
FT /label= M18
FT Peptide 121..135
FT /label= M12
XX
XX W09A06421-A.
XX
XX 31-MAR-1994.
XX
XX 15-SEP-1993; 93WO-US08703.
XX PF
XX 16-SEP-1992; 92US-0945954.
XX PR
XX (UYTE-) UNIV TENNESSEE RES CORP.
XX PA
XX

PI Dale JB, Lederer JW;
XX
XX WPI; 1994-118122/14.
DR N-PSDB; Q45223.
XX
XX New immunogenic hybrid proteins derives from streptococcal M proteins
PT - induces opsonic antibodies, for protective immunisation against
PT against multiple group A streptococci serotypes
XX
XX Disclosure: Fig 13; 67pp; English.
XX
XX The sequences given in R50992-1001 represent hybrid M proteins which
CC contain the M24-M5-M6 and/or M19 subunits. These multivalent
CC proteins were constructed using fragments of the 5' regions of emm
CC genes that were amplified by PCR, ligated in tandem and expressed in
CC PKK223.3. The amplified regions pref. encode protective and not
CC tissue-cross-reactive epitopes, which can then be linked into one
CC protein molecule. The recombinant hybrid protein may contain 113
CC N-terminal amino acids of M24, 58 amino acids of M5, 35 from M6 and
CC 35 from M19. Each section is linked by 2 amino acids specified by
CC the respective restriction enzyme sites that were synthesised into
CC the primers used to specify the PCR product. Multivalent M proteins
CC such as this may be used for protective immunisation against group A
CC streptococci, which esp. cause rheumatic fever and rheumatic heart
CC disease. Humoral antibodies raised against these proteins do not
CC react with heart tissue antigens but are effective against many
CC different serotypes. The multivalent proteins may also include
CC sequences which induce mucosal antibodies and do not require coupling
CC to an immunogenic carrier.
XX
XX Sequence 135 AA;

Query Match 17.3%; Score 338; DB 15; Length 135;
Best Local Similarity 37.4%; Pred. No. 5e-16;
Matches 92; Conservative 7; Mismatches 23; Indels 124; Gaps 6;

QY 3 MVATRSQTDTLEKQVQRADKFEIENNTLKNSDLSFNKALKDHNDLTELSELSNAKEKL 62
DB 1 MVATRSQTDTLEKQV-----
QY 63 RGSATVTRGTINDPQRAKEALDKYLENHDLTKNEGLKTENEGLKTENEGLKT 122
DB 17 -gsavtrgtindpqrak-----
QY 123 EVDVFPRTGTVENPDKARELLNKYDVENSMLQANDKLPWRVRYTRHTPEDKIKKIIDDL 182
DB 33 evdrvfpgrtvenpdkar-----pwrvrytrhtpedkikl---q 69
QY 183 DAKHEHLOQONEKLSLQNGDGNPREVIEDLAANNPAIONIRLRHENKD-----L 231
DB 70 darsvngfprhvkldngdgnpreviedlaaefapi--tratadnkdelirshsdvae 127
QY 232 KARLEN 237
DB 128 kqried 133

RESULT 15
B03124
ID B03124 standard; Protein; 135 AA.
XX B03124;
XX
XX 10-OCT-2000 (first entry)
XX
XX S. pyogenes octavalent hybrid M protein, SEQ ID NO:22.
XX Multivalent hybrid M protein; group A streptococcus; serotype;
KW immunogenic; sero-specific antibody; streptococcal infection;
KW cross reactivity; vaccine; acute rheumatic fever; ARF;
KW rheumatic heart disease; streptococcal pharyngitis; strep throat;
KW pneumonia.

XX Streptococcus pyogenes.
OS Synthetic.
XX US6063386-A.
PD 16-MAY-2000.
XX 15-SEP-1997; 97US-0937271.
XX 16-SEP-1992; 92US-0945954.
PR (UYTE-) UNIV TENNESSEE RES CORP.
XX Lederer JW, Dale JB;
PI WPI; 2000-364475/31.
DR N-PSDB; A57902.
XX
PT New immunogenic recombinant hybrid M protein comprising amino-terminal
PT peptide fragments of streptococcal M protein useful as vaccine against
PT rheumatic fever and infections leading to rheumatic fever
XX
PS Disclosure; Fig 13; 62pp; English.
XX

CC The invention relates to multivalent immunogenic hybrid group A
CC streptococcal M proteins comprising N-terminal peptide fragments of M
CC proteins that elicit opsonic antibodies against multiple serotypes
CC of group A streptococci (e.g., Streptococcus pyogenes). The antibodies
CC generated using the hybrid proteins are against one or more M protein
CC serotypes, where at least one of the serotypes is M1, M3, M18 or M19.
CC The invention also encompasses a recombinant DNA molecule comprising a
CC nucleotide sequence that encodes a multivalent hybrid M protein; and a
CC method for immunising a mammal against streptococcal infections,
CC comprising administering an immunogenic multivalent hybrid M protein to
CC the mammal. The multivalent hybrid M proteins are useful for eliciting
CC opsonic or protective antibodies to the M proteins of different serotypes
CC of group A streptococci, and may therefore be used as vaccines to protect
CC against and control infection by type A streptococci. Type A streptococci
CC are not only responsible for streptococcal pharyngitis (strep throat),
CC forms of pneumonia and a condition resembling toxic shock, but are also
CC involved in the development of acute rheumatic fever (ARF) and rheumatic
CC heart disease. In a patient with ARF, antibodies formed during a group A
CC streptococcal infection are also cross-reactive with heart tissue, which
CC indicates that the streptococci and host tissue contain similar antigenic
CC motifs. The new multivalent vaccines are capable of raising sero-specific
CC antibodies against various serotypes of group A streptococci which are
CC not cross-reactive with human heart tissue. Sequences B03113-B03117,
CC B03119-B03121 and B03123-B03124 represent multivalent hybrid
CC Streptococcus pyogenes M proteins generated in the disclosure of the
CC invention.
XX
SQ Sequence 135 AA;

Query Match 17.3%; Score 337.5; DB 21; Length 135;
Best Local Similarity 38.6%; Pred. No. 5.4e-16;
Matches 88; Conservative 6; Mismatches 21; Indels 113; Gaps 5;

QY 3 MVATRSQDTTLEKVOERADKFEIENNTLKNLSLSPNNKALDKHNDDELTELSNAKEKL 62
|||||
Db 1 mvatrsqdttlekvqe----- 16

QY 63 RGSVTRGTINDPQAKALDKYLELNDHDKTKNEGLKTENEGLKTENEGLKT 122
|||||
Db 17 -gsavtrgtindpqrak----- 32

QY 123 EYDRVFPRGTVENPKARELLNKYDVNSMLQANNDKLPWRVYRTRHPEDKLLKIIDDL 182
|||||
Db 33 evdrvfprgtvenpkar-----pwrvytrtpeklkl---q 69

QY 183 DAKHELOQNEKLSLQDGNPREVIEDLAANNPAIQNIRLHENKD 230
||: : ||||| : : |||

Db 70 darsvngfeprhvklihdngdgnpreviedlaaefapl--trataadnkd 115

Search completed: May 10, 2001, 08:38:14
Job time: 67 sec

QY 244 RDKFRAGTLLDQVTLTKHNSNYQYNAQAQRDLRQKAEYKGLNDWAERLLQELNID 303
Db 136 ----- 135
QY 304 VATRSQTDLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKLR 363
Db 136 ----- 135
QY 304 VATRSQTDLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKLR 363
Db 136 ----- 135
QY 364 KNDKSLSEKASKIOLEARK 383
Db 196 KNDKSLSEKASKNOLEARK 215
RESULT 4
US-08-937-271-17
; Sequence 17, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 343 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-17

Query Match 31.8%; Score 622; DB 3; Length 343;
Best Local Similarity 39.2%; Pred. No. 1.7e-38;
Matches 149; Conservative 10; Mismatches 17; Indels 204; Gaps 3;
QY 4 VATRSQTDLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKLR 63
Db 40 VPRGTVENPDKARELLNKYDVNSMLQANDNLV----- 74
QY 64 GSAVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTENEGKLTENEGKLTKE 123
Db 75 -DAVTRGTINDPQRAKEALDKYELENHDLTKNEGLKTENEGKLTENEGKLTKE 133
QY 124 VDRVPRGTVENPDKARELLNKYDVNSMLQANDNLKPWRVRYRTRHPEDKLLKIIDLD 183
Db 134 -----PW----- 135
QY 184 AKEHELOQONEKLSLQDGNPREVIEDLAANPAIONIRLHRHENKDKIARLENAMVAG 243

Db 136 ----- 135
QY 244 RDKFRAGTLLDQVTLTKHNSNYQYNAQAQRDLRQKAEYKGLNDWAERLLQELNID 303
Db 136 ----- 135
QY 304 VATRSQTDLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKLR 363
Db 136 ----- 135
QY 364 KNDKSLSEKASKIOLEARK 383
Db 196 KNDKSLSEKASKNOLEARK 215
RESULT 5
US-08-937-271-15
; Sequence 15, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 173 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-15

Query Match 29.5%; Score 576; DB 3; Length 173;
Best Local Similarity 69.9%; Pred. No. 1.8e-35;
Matches 121; Conservative 0; Mismatches 0; Indels 52; Gaps 1;
QY 3 MWATRSQTDLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKL 62
Db 1 MWATRSQTDLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLDELTEELSNAKEKL 60
QY 63 R-----GSAVTRG 70
Db 61 RKNDKSLSEKASKNOLEARKADLEKALEGAMNFSTADSAKITLEAKADLEGSAVTRG 120
QY 71 TINDPQRAKEALDKYELENHDLTKNEGLKTENEGKLTENEGKLTKE 123
Db 121 TINDPQRAKEALDKYELENHDLTKNEGLKTENEGKLTENEGKLTKE 173

RESULT 6
US-08-937-271-2
; Sequence 2, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-2

Query Match 20.1%; Score 392; DB 3; Length 287;
Best Local Similarity 100.0%; Pred. No. 9.7e-22;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 304 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKL 363
Db 2 VATRSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKL 61

Qy 364 KNDKSLSEKASKIOELPARK 383
Db 62 KNDKSLSEKASKIOELPARK 81

RESULT 7
US-08-937-271-22
; Sequence 22, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/937,271
FILING DATE: 15-SEP-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman, Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 481112.405C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-937-271-22

Query Match 17.3%; Score 337.5; DB 3; Length 135;
Best Local Similarity 38.6%; Pred. No. 3.8e-18;
Matches 88; Conservative 6; Mismatches 21; Indels 113; Gaps 5;
Qy 3 MVATRSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELSELSNAKEKL 62
Db 1 MVATRSQDTLEKVOE----- 16
Qy 63 RGSVTRGTINDPQRAKEALDKYELEHDLTKYNEGLKTENEGLKTENEGLKT 122
Db 17 -GSVTRGTINDPQRAK----- 32
Qy 123 EYDVPFPGTGVENPDKARELLNKKYDVNSMLQANNDKLPWRYVTRHTPEDKLLKIIDL 182
Db 33 EYDVPFPGTGVENPDKAR-----PWRYVTRHTPEDKLLKL---Q 69
Qy 183 DAKEHELOQNEKLSLQNGDGNPREVIEDLAANNPAIQNIRLRHENKD 230
Db 70 DARSVNGEPFHVHVKLIDNGDGNPREVIEDLAAEFAPL--TRATADNKD 115

RESULT 8
US-08-937-271-10
; Sequence 10, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:

; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-10

Query Match 17.1%; Score 335; DB 3; Length 305;
Best Local Similarity 29.7%; Pred. No. 1.6e-17;
Matches 115; Conservative 30; Mismatches 70; Indels 172; Gaps 12;
QY 3 MVATRSQDTTLEKVOERADKFEIENNTLKNSDLSFNKALKDHNDLTELSENAKEKL 62
Db 1 MVATRSQDTTLEKVOE-----16
QY 63 RGSVTRGTINDPQRAKALDKYELENHDLTKNEGLKTENEGLKTENEGLKT 122
Db 17 -GSVTRGTINDPQRAK-----32
QY 123 EVDVFRPGTGVENPKARELLNKYDVNSMLQANNDKLPWRVRYTRHPEDKLLKIID-- 180
Db 33 EVDVFRPGTGVENPKAR-----PWRVRYTRHPEDKLLKLQNKI 72
QY 181 -----DLDAK---EHELQOONEKLSLON--GDGNPREVIEDLAANNPAIQNIRL 225
Db 73 SDASRKGRLRDLDASREAKKQLEAGHQLEQNKISEASRKGRLRDLDAS-----R 123
QY 226 HENKDKARLENAMVAGRDFKFRAGTLLDQVTLTKHNSYQYNAQAGRLDLRQKAEY 285
Db 124 EAKKQLEAEQKLE-----QNKISEASRKGRLRDLD 156
QY 286 LKGLNDWAERLQELNIDVATRSQDTTLEKVOERADKFEIENNTLKNSDLSFNKALK 344
Db 157 SREAKQVEKALEANSKLAA---LEKLNKELESKLTKEKAELOAK---LEAEAKAL 210
QY 345 KHNDELTELSNAKEKLRKNDKSLSE 371
Db 211 KEQLAKQAEELA----KLKAGKASDSQ 233

RESULT 9
US-08-937-271-8
; Sequence 8, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937/271
; FILING DATE: 15-SEP-1997

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-8

Query Match 16.4%; Score 321; DB 3; Length 187;
Best Local Similarity 45.2%; Pred. No. 9.1e-17;
Matches 84; Conservative 19; Mismatches 43; Indels 40; Gaps 7;
QY 3 MVATRSQDTTLEKVOERADKFEIENNTLKNSDLSFNKALKDHNDLTELSENAKEKL 62
Db 1 MVATRSQDTTLEKVOEVATRSQDT-----TLEKVOEVATRSQDTTLEKV 44
QY 63 R-GSAVTRGTINDPQRAKALDKYELENHDLTKNEGLKTENEGLKTENEGLK 121
Db 45 QEGSAVTRGTINDPQRAKAEAVTRGTI--NDPQRAKEAV-----TRGTINDPQRAK 92
QY 122 TEVDVFRPGTGVENPKARELLNKYDVEN-----SMLQANNDKL-PWRVRYTRHTP 171
Db 93 -EVDVFRPGTGVENPKARVFPRTGVENPDKARVFPRTGVENPDKARVRYTRHTP 151
QY 172 EDKLLK 177
Db 152 EDKLLK 157

RESULT 10
US-08-795-475-6
; Sequence 6, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bivrock, Lars
; APPLICANT: Sjvbring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795/475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 443 amino acids

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-475-6

Query Match 16.0%; Score 312; DB 2; Length 443;
Best Local Similarity 43.1%; Pred. No. 1.2e-15;
Matches 93; Conservative 18; Mismatches 65; Indels 40; Gaps 7;
QY 200 NCDGNPREVIEDLAANNPAIONIRLRHENKDKARLENAMVAGRDFKRA-----GTL 252
DB 1 NGDGNPREVIEDLAANNPAIONIRLRHENKDKARLENAMVAGRDFKRAEKEKAKQAL 60
QY 253 LDQVLTQYTKHNSVQOY-----NAQAGRLDRLQKAEYLKGLNDWAER----- 295
DB 61 EDQRDLTKLKELOQDYDLAKESTSWDRQRLEKELEKEKALELAIDQASRDYHRATAL 120
QY 296 -----LLOELNIDVATR--SQTDLKVKQVRADK--FEIENNTL---KLKNSDLSFNN 341
DB 121 EKELEKKKALELAIDQASQDYNRANVLEKELETITREQEINRNLGNKALELDQLSSEK 180
QY 342 KALKDHNDLTFEE--LSNAKEKLRKDKSLSEKASK 375
DB 181 EOLTIKAEKLEEKQISDASQSRRLDLDASREAKK 216

RESULT 11
US-08-937-271-20
; Sequence 20, Application US/08937271
; Patent No. 6063386
; GENERAL INFORMATION:
; APPLICANT: Dale, James B.
; APPLICANT: Lederer, James W.
; TITLE OF INVENTION: RECOMBINANT MULTIVALENT M PROTEIN
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,271
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman, Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 481112.405C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 67 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-937-271-20

Query Match 13.7%; Score 267; DB 3; Length 67;
Best Local Similarity 38.3%; Pred. No. 2.3e-13;
Matches 67; Conservative 0; Mismatches 0; Indels 108; Gaps 3;

QY 3 MVATRSQTDITLEKVOVRADKFEIENNTLKLKNSDLSFNNKALKDHNDLTELSELSNAKEKL 62
DB 1 MVATRSQTDITLEKVOE----- 16
QY 63 RGSAVTRGTINDPQRAKEALDKYLELHDLTKNEGLKTENEGLKTEENGLKTENEGLKT 122
DB 17 -GSAVTRGTINDPQRAK----- 32
QY 123 EYDVRFPRTGVENPDKARELLNKKYDVNSMLQANNDKLPWRVRYTRHTPTEDKLKK 177
DB 33 EYDVRFPRTGVENPDKAR-----PWRVRYTRHTPTEDKLKK 67
RESULT 12
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 9.6%; Score 188.5; DB 1; Length 2482;
Best Local Similarity 21.7%; Pred. No. 1.1e-05;
Matches 110; Conservative 92; Mismatches 145; Indels 159; Gaps 22;

QY 8 SQTDLKVKQVRADKFEIENNTLKLKNSDLSFNN--KALKD---HNDELTELSELSNAKEKL 62
DB 1424 NQIAQINKEKELLVK-ESESQARLSESDYEKLNVSKALEALVKEGFALRLSTQGEV 1482
QY 63 -----RGSAVTRGTINDPQRAK-----EALDKYLELHDLTKNEGLKTENEGLKTEENGL- 113
DB 1483 HOLRRGIEKLRVRIEADKQKHIAEKLKERERENDSLKDKVENLERELQ-MSEENQELV 1541
QY 114 -----KTENEGLKTEVDVRFPRGTVENPKARELLNKKYDVNSMLQANNDKLPWRV-- 164
DB 1542 ILDAENSKAEVETLKTQIE-----EMARS-LKVFELDLVTLRSEKENLTQIQE 1589


```
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match          9.6%; Score 187.5; DB 5; Length 3248;
Best Local Similarity 21.7%; Pred. No. 1.8e-05;
Matches 110; Conservative 92; Mismatches 145; Indels 159; Gaps 22;

QY 8 SQTDTLEKVRADKFEIENNTLKNKSDLSFNN--KALKD---HNDELTELSNAEKL 62
DB 2152 NQIAQLNKEKELLVK--ESELQARLSSESDYEKLNVSKALEAALVEKGFEALRLSSTQEEV 2210
QY 63 ---RGSVTRGTINDPORAK---EALDKYLENHDLTKNKNEGLKTENEGKL- 113
DB 2211 HOLRGRGIEKLVRIEADKKHIAEKLKERENDSLKDKVENLERELQ-MSEENGELV 2269
QY 114 -----KTENEGKTEVDVPRGTVENPKARELLNKYDVNSMLQANNDKLPWRV-- 164
DB 2270 ILDAENSKAEVETLKTQIE-----EMARS-LKIFELDLVTLRSSEKENTKQIQE 2317
QY 165 RYTRHTPEDKL---KKIIDDLDAKE-----HELOQNEKLSLQONGDGNP 205
DB 2318 KQGQLSELDKLSLSPKSLKEEQAEIQIKESKTAVEMLQNLQKELNEAVALCGD--- 2374
QY 206 REVIE--DLAANNPAIQNLRLHENKDLKARLENAMVAGDFKFRAGTLLDDVDTQLYTKH 263
DB 2375 QEIMKATEQSDDPPIEEBHQRLNSTEKURALE-----ADERKKQVCVLOQLKE--SEH 2425
QY 264 NSNY-----QQYNAQAGRLDLQKAEYKGLNDWAERLIQELNIDVAT 306
DB 2426 HADLLKGRVENLERLEARTNOEHAALAEANSKGEVETLAKIEGTQSLRGLDELVDVT 2485
QY 307 -RSQ-----TDLEKQVORADKFEIENNT-----LKLKNSDL-----S 338
DB 2486 IRSEKENLTNLOKEQERISELEIINSFENILOKEQEQVMKEKSTAMEMLTQOLKE 2545
QY 339 FNNKALKOHND------LTEE 354
DB 2546 LNERVAALHNDQEAQKAEQNLSSQVECLEKALQLOGLDEAKNNYIVLOSSVKGLIQE 2605
QY 355 LSNAKEKLRKNDKSLSEKASKIQELE 380
DB 2606 VEDGKQKLEKDEISRLKNQIQDQE 2631

RESULT 15
US-08-630-822A-62
; Sequence 62, Application us/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDIA
; TITLE OF INVENTION: NOVEL ECOTOPARASITE SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C. Suite 3500
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,822A
; FILING DATE: 11-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-9223
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 900 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-630-822A-62

Query Match          9.5%; Score 185; DB 2; Length 900;
Best Local Similarity 24.2%; Pred. No. 5.8e-06;
Matches 104; Conservative 70; Mismatches 175; Indels 80; Gaps 19;

QY 7 RSQDTFLK-----VQERADKFEIENNTLKNKNSDLSPNNKALKDHNDELTELSNA 58
DB 423 RNQAGDYKELTKFKLSCKELQKAEFELNRKRTESLLVETKKRLDEQNKRTREMN 482
QY 59 KEKLRGSVTRGTINDPQRAKALDKYLENHDLTKNKGLK-----TE--NEGLKTE 109
DB 483 QOHNDKINLEKQINDLQ---EKL-KGELE-HNOKLKKQAVELRVAQSATEQLNNELOET 537
QY 110 NEGLKTENEGKTEVDVPRGTVENPKARELLNKYDVNSMLQANNDKLPWRVYTRH 169
DB 538 MQGLQTRDALQOEVASL--QGLSQSRSRSQASDMQIE---LEAKLOALHIELEHVRN 592
QY 170 TPEDKLLKIIDDLDAKEHELOQNEKLSLQONGDGNPREVIEDLAANNPAIQNLRLHENK 229
DB 593 C-EDKVTQ--DNQQLLERISTLEKECASLE-----LELKATQNKYEQEVKAHRETE 640
QY 230 DLKARLENAMVAGDFKFRAGTLLDDVDTQLYTKHNSNYQQYNAQAGRL-----277
DB 641 --KSRLVSKEEANMEVKALQIKLNEEKSAKQSDQNSQEKERQISMLSVDYRQIQORLQ 698
QY 278 ----DLRQKAEYKGLNDWAEL-----RLQLNELNIDVATRSQDTL-----EKVQ 317
DB 699 KLEGEYRQSEKVKALHSQIEQEQQLKKSQIQSELGVQ---RSQTAHLTAREAQLVGEVAH 755
QY 318 ERADKFEIENNTLKNKNSDLSPNNKALKDHNDEL-TEELSSNAKEKLRKND--KSLSEKAS 374
DB 756 LRDAKRNVEEELHKLTA-RSVDNAQMKLEQEQVEAEQVFSTLYKTHSNELKEEKS 814
QY 375 KIQELEARK 383
DB 815 HIQEMEER 823

Search completed: May 10, 2001, 08:37:51
Job time: 44 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:07 ; Search time 18.66 seconds
(without alignments)
1410.551 Million cell updates/sec

Title: US-09-151-409-16

Perfect score: 1954

Sequence: 1 ACWATRSQDTLEKVOERA.....KNKSLSEKAKIOLEARK 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_67:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	397	20.3	539	2	A28549
2	385.5	19.7	492	2	A28616
3	339	17.3	197	1	MMSOMP
4	312	16.0	484	2	S46489
5	311	15.9	484	2	S35401
6	302	15.5	227	2	S01260
7	257.5	13.2	483	2	A26297
8	252	12.9	83	2	S61086
9	248	12.7	484	2	S34978
10	227	11.6	2139	2	T18296
11	220.5	11.3	327	2	F71980
12	214.5	11.0	150	2	S60838
13	212.5	10.9	1526	2	T41522
14	207.5	10.6	980	2	E71606
15	206.5	10.6	1931	2	A59234
16	206.5	10.6	2094	2	S33124
17	205	10.5	1392	2	A43336
18	204.5	10.5	520	2	S35575
19	204.5	10.5	911	2	S51441
20	204.5	10.5	1927	2	A59236
21	203	10.4	1827	2	T16270
22	203	10.4	1992	1	S02771
23	202	10.3	1427	2	S22695
24	200	10.2	1208	2	T39068
25	200	10.2	1939	1	A46762
26	199.5	10.2	1738	2	T14867
27	199	10.2	1509	1	A27224
28	198.5	10.2	2954	2	T14156
29	197	10.1	764	2	I51302

30 197 10.1 1496 2 T05634
31 197 10.1 1837 2 T41023
32 197 10.1 1939 2 T48175
33 196 10.0 1979 2 C71622
34 195.5 10.0 2288 2 T29999
35 195 10.0 1937 2 T38055
36 195 10.0 1938 2 T49464
37 193.5 9.9 1875 2 S38173
38 193.5 9.9 1934 2 T48153
39 193.5 9.9 1935 2 A59286
40 193.5 9.9 1957 2 T38077
41 193 9.9 875 2 S70115
42 193 9.9 1790 2 S67593
43 192.5 9.9 1935 1 A37102
44 192.5 9.9 1935 1 S06006
45 192 9.8 98 2 S61079

hypothetical prote
probable nuclear p
myosin heavy chain
hypothetical prote
hypothetical prote
myosin heavy chain
alpha cardiac myos
myosin-like protei
myosin heavy chain
hypothetical coile
ZiP1 protein - yea
transport protein
myosin beta heavy
myosin beta heavy
M protein precuro

ALIGNMENTS

RESULT 1

A28549

M24 protein precursor - Streptococcus pyogenes

A:Species: Streptococcus pyogenes

A:Variety: serotype M24

C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 26-Aug-1999

C:Accession: A28549; S60802

R:Mouw, A.R.; Beachey, E.H.; Burdett, V.

J. Bacteriol. 170, 676-684, 1988

A:Title: Molecular evolution of streptococcal M protein: cloning and nucleotide sequ

A:Reference number: A28549; MUID:88115166

A:Accession: A28549

A:Molecule type: DNA

A:Residues: 1-539 <MOD>

A:Cross-references: GB:M19031; NID:g153616; PIDN:AAA26874.1; PID:g153617

R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.

Mol. Microbiol. 14, 619-631, 1994

A:Title: Non-congruent relationships between variation in emm gene sequences and the

A:Reference number: S60784; MUID:95198537

A:Accession: S60802

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual

A:Molecule type: DNA

A:Residues: 30-89 <NHA>

C:Superfamily: M5 protein

C:Keywords: coiled coil; transmembrane protein

Query Match 20.3%; Score 397; DB 2; Length 539;

Best Local Similarity 33.0%; Pred. No. 4 4e-12;

Matches 133; Conservative 57; Mismatches 155; Indels 58; Gaps 13;

QY 4 VATRSQDTTLEKVOERADKFEIENNTLKNSDLSFNKALKDHNDELTEELSENAKEKL 63

Db 43 VATRSQDTTLEKVOERADKFEIENNTLKNSDLSFNKALKDHNDELTEELSENAKEKL 102

QY 64 GSAVTRGTINPQRAKALDKYLE--NHDLKTKNEG-----LKTENEGKLTEN 110

Db 103 K-----NDKSLSEKASKIQLEAKADKALEGAMNFSTADSAKIKT-----LEAEK 150

QY 111 EGLKTENEGKLTEDVRFPRGTVPENPKARELLNKYDVNSMLQANNDKLPWRVRYTRH- 169

Db 151 AALAARKADLEKALEGAMNFSTA--DSAK--IKTLEAKAALAEKALEKALEGAMNF 205

QY 170 --TPEKDKLKIIDD---LDAKEHELOQONE-KUSLONGDNPREVIEDLAANNPAIQNR 223

Db 206 STADSASIKIKTLEAKAALAEKADKALEGAMNFSTADSAKIKTLE-----AEKAA 257

QY 224 LRHENKDKLARENAMVAGRDVFKRAGTILDDQVTLTYT-----KHSNTYQQVNAQGRLD 278

Db 258 LEARQAELEKALEGAMNFSTADSAKIKTLEAKAALAEKADKALEKALEKALEK 317

QY 279 LRQAEYLKGLNDWAERLLQELNIDVATR-SQTDLTLEKVOERADKFEIENNTLKNSDL 337

MMSOMP
 M5 protein - Streptococcus pyogenes (fragment)
 C.Species: Streptococcus pyogenes
 C.Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 16-Feb-1997
 C.Accession: A03501
 R:R.Manjula, B.N.; Acharya, A.S.; Mische, S.M.; Fairwell, T.; Fischetti, V.A.
 J. Biol. Chem. 259, 3686-3693, 1984
 A.Title: The complete amino acid sequence of a biologically active 197-residue fragment
 A.Reference number: A03501; MUID:84162039
 A.Accession: A03501
 A.Molecule type: protein
 A.Residues: 1-197 <MAX>
 C.Comment: Residues 27-58 contain several tandem 7-residue repeats, and residues 150-160 contain a 10-residue repeat.
 C.Comment: This protein is closely associated with the virulence of the bacterium and is a major virulence factor.
 C.Superfamily: M5 protein
 C.Keywords: cell wall; transmembrane protein; virulence

	Query Match	17.3%;	Score 339;	DB 1;	Length 197;
	Best Local Similarity	44.2%;	Pred. No. 7.5e-10;		
	Matches 87;	Conservative 23;	Mismatches 49;	Indels 38;	Gaps 7;
Qy	67	VTGTTNDPQRAEADKYELEHNHDKLTNEGKLTNEGKLTNEGKLTNEGKLTVEVD	126		
		: : : : : : : : : :			
Db	2	VTGTTSDPQRAEADKYELEHNHDKLTNEGKLTNEGKLTNEGKLTNEGKLT---	58		
		: : : : : : : : :			
Qy	127	VFRGTGVNPDKRLNLYDVNSMLQANNDKLPWRVRYTRHTPDKLKIKIDDLAKE	186		
		: :			
Db	59	---KSNLIER--KYAELTS-----EKKEHEAENDKLK--QORDTLSTQKETLEREVQ	103		
		: :			
Qy	187	HELQQONEKLSLQNGD-----GNPREVIEDLAANNPAIQ-----NIRURHEN	228		
		: :			
Db	104	---GVNNFTKINGNDTKELNKTBFQIANKQOESFNKELNELLKTKTVKDKIAKEOE	161		
		: :			

RESULT 4
S46489
M1 protein precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C:date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-May-1999
C:Accession: S46489; S46490
R:Akesson, P.; Schmidt, K.H.; Cooney, J.; Bjoerck, L.
Biochem. J. 300, 877-886, 1994
A:title: M1 protein and protein H: IGFIC- and albumin-binding streptococcal surface p
A:reference number: S46489; MUID:94280417
A:Accession: S46489
A:Molecule type: DNA
A:Residues: 1-484 <AKE>
A:Experimental source: strain 40/58, serotype M1
A:Accession: S46490
A:Molecule type: protein
A:Residues: 42-51 <AKW>
A:Experimental source: strain 40/58, serotype M1
C:Genetics:
A:Gene: emm1
C:Superfamily: M5 protein
C:Keywords: transmembrane protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-484/Product: M1 protein #status experimental <MAT>
F:459-477/Domain: transmembrane #status predicted <TM>

```

Query Match      16.08;  Score 312;  DB 2;  Length 484;
Best Local Similarity 43.18;  Pred. No. 3.9e-08;
Matches 93;  Conservative 18;  Mismatches 65;  Indels 40;  Gaps 7;

Qy 200  NGDGNPREVIEDLAANNPAIQIRLRHKNKDLKARLENAMEVAGRDFKRA-----GTL 252
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 42  NGDGNPREVIEDLAANNPAIQIRLRHKNKDLKARLENAMEVAGRDFKRAELEKQKAL 101

```

```

Query Match      15.9%; Score 311; DB 2; Length 484;
Best Local Similarity 44.2%; Pred. No. 4.4e-08;
Matches 96; Conservative 22; Mismatches 61; Indels 38; Gaps 19;
```

QY	200	NGDGNPREVIEDLAANNPAIONIRLRHENKDKLKALENAMAVAGRDFKRA-----GTL	252
Dd	42	NGDGNPREVIEDLAANNPAIONIRLRHENKDKLKALENAMAVAGRDFKRAEELEAKAQAL	101
QY	253	LQVTQLTKTHNSNYQQYN--AAQGRDLRLQRKAEYLKLNDWAERLLQEINLDIVATRS--Q	309
Dd	102	EQRKDLETKLKLQQDYDLAKESTSDNRQLE--KELEEKKCAL--ELAIQAASRDYHR	157
QY	310	TDTLEKVQE-----RADKFIEIENNTLK---NSDSLFFNNKALKDNH	348
Dd	158	ATAPELEKEEKKALELAIDQASODYRRANVLYKEKELDTITREQEINRNLIIGNRKLELDQL	217
QY	349	DELTEELSNAEKLRKNDSLSSE--KASKIQELEARK	383
Dd	218	SSEKQOLTIETIAKL--EEKQISDASROSLRRDDAASR	253

RESULT 6

Query Match	13.2%;	Score 257.5;	DB 2;	Length 483;
Best Local Similarity	29.0%;	Pred. No. 1.4e-05;		
Matches 105; Conservative	40;	Mismatches 134;	Indels 83;	Gaps 13;

Query Match	13.2%;	Score 257.5;	DB 2;
Best Local Similarity	29.0%;	pred. No. 1.4e-05;	
Matches 105; Conservative	40;	Mismatches 134;	Indels 83

Db 1426 YEKVLKRKADLEAEANENLDIEKKDRMKKQVKKLEGELKETKDKLNAIAAEKDSIFTA 1485
Qy 130 -----RGVYENPKARELLNKYDVNSMLQANNKLPWRVRYTRHTPEDKLKKI 178
Db 1486 KQSDADLEELNKTVESHD---EVVALNTQITKLTRDNQSAEEELNELR-SKADKDKK 1541
Qy 179 IDDLDAKEHELOQOONKLSLQNGDGNPREV-----IEDL--AANPAIQNIRLRHENKDLK 232
Db 1542 ISELEEQVNEL--ESRPVGTGNADENEIKIRDAQIADLNKALEMKGVQNNQLQATNKELK 1599
Qy 233 A-----RLNAME-----VAGRDFKRACTLLDQVQTYLYTKH 263
Db 1600 AKNDLTSKIBITENEMKLENKAKRLEQDKDEADKAVSQTIKRGK-LEEEVKKLTTEI 1658
Qy 264 NSNYQOYNAQAGRLDLRQAEYKLGNDWAERLLLOELNIDVATRSQTDITLEKVOERAD-- 321
Db 1659 QALFKQINAPSSVAQEERKORLESIDIAELKEQLEQERT--TAANAER-KKIQAELEDEV 1715
Qy 322 KFEIENNT-----LKKNSD-----LSFNKALKDH-----NDELTEELSNAREK 361
Db 1716 KFNLEDVYNQREKLVAKNSENDAEIDSLEKEKALEDEIEKITDDNNKLEESIDSLDRKY 1775
Qy 362 ---LRKNDKSLSEKASIQLEARK 383
Db 1776 NALLDSKSDSVSMKEKQDELKVTK 1800
RESULT 11
F71980
hypothetical protein jhp0052 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: F71980
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: F71980
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <ARN>
A:Cross-references: GB:AE001444; GB:AE001439; NID:g4154549; PIDN:AA05629.1; PID:g415455
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0052

Query Match 11.3%; Score 220.5; DB 2; Length 327;
Best Local Similarity 25.7%; Pred. No. 0.00051;
Matches 97; Conservative 64; Mismatches 133; Indels 83; Gaps 15;
Qy 14 EKVOERADKFEIENNTLKNLSLQNGDGNPREV-----IEDL--AANPAIQNIRLRHENKDLK 1599
Db 16 EELARISELEDENTELLREYLAETSELKDANDQLRQKNDKLFITDKL----- 67
Qy 71 TINDPQRAKALDYELNHDLYKNEGLKTENEGLKTENEGLKTENEGLKTENEGLKTENEGLKT 130
Db 68 -----TKNTELF-AENESLSVKISGLEHSDQLWQNNKLTKEKALTE----- 112
Qy 131 GTVENPKARELLNKYDVNSMLQANNKLPWRVRYTRHTPEDKLKIIDDLDAKEHELOQ 190
Db 113 -----KDILAK-----ENRLLAARDRLTEKRELT-TEKERLARENTELTHTKLTET 159
Qy 191 QONKLSLQNGDGNPREVIEDLAANPAIQNIRLRHENKDLKARLENAMVAGRDFKFRAG 250
Db 160 KENKALATEN-----DKLNHQVTALTNERDSLEQ-----RARLQDA-----HG 198
Qy 251 TLLDQVQTYLYTKHNSYQOYNAQAGRLDLRQAEYKLGNDWAERLLLOELNIDVATRSQT 310
Db 199 FLEKRCNTNL-EKEN---ORUTDKLKQLESQAOKS--LENTNNQLQALENSVLOAQAKEK 252

Qy 311 DTLEKVOERADKFEIENNTLKN-----SDLSFNKALKOHNDLTELTEELSNAREK-- 361
Db 253 IAIE-----KSELERETARLUKSLGMEAKSDLDLHNRRLASANEKLRQNRKLEENI 305
Qy 362 -LRKNDKSLSEKASKIQ 377
Db 306 ALKERVLDGLNEQLSKIQ 322
RESULT 12
S60838
M protein precursor - Streptococcus pyogenes (serotype M6) (fragment)
C:Species: Streptococcus pyogenes
A:Variety: serotype M6
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C:Accession: S60838
R:Whitmore, A.M.; Kapur, V.; Sullivan, D.J.; Musser, J.M.; Kehoe, M.A.
Mol. Microbiol. 14, 619-631, 1994
A:Title: Non-congruent relationships between variation in emm gene sequences and the
A:Reference number: S60784; MUID:95198537
A:Accession: S60838
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-150 <WHA>
A:Cross-references: EMBL:U11986; NID:g533649; PIDN:AAA99602.1; PID:g1235835
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Superfamily: M5 protein

Query Match 11.0%; Score 214.5; DB 2; Length 150;
Best Local Similarity 43.6%; Pred. No. 0.0004;
Matches 61; Conservative 11; Mismatches 25; Indels 43; Gaps 6;
Qy 126 RVFPRTGVENPKARELLNKYDVNSMLQANNKLPWRVRYTRHTPEDKLKIIDDLDAK 185
Db 30 RVFPRTGVENPKARELLNKYDVNSMLQANNKLT-----TTENK----- 69
Qy 186 EHELOQOONKLSLQNGDGNPREVIEDLAANPAIQNIRLRHENKDLKARLENAMVAGRD 245
Db 70 ---NLTDQNKLDQNN-----KELKA-----EENRLTTENKGLTKKLSAEAEAAANK 113
Qy 246 FKRA-----GTL---LDQVQ 258
Db 114 EQESKETIGTLAKTLDEIVK 133

RESULT 13
T41522
myosin ii - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000
C:Accession: T41522
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z22000
A:Accession: T41522
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1526 <WOO>
A:Cross-references: EMBL:AL049498; PIDN:CAB39901.1; GSPDB:GN00068; SPDB:SPCC645.05C
A:Experimental source: strain 972h-; cosmid c645
C:Genetics:
A:Gene: SPDB:SPCC645.05C
A:Map position: 3
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:80-743/Domain: myosin motor domain homology <MCO>

Query Match 10.9%; Score 212.5; DB 2; Length 1526;
Best Local Similarity 22.1%; Pred. No. 0.007;
Matches 93; Conservative 74; Mismatches 161; Indels 93; Gaps 12;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:52 ; Search time 11.75 Seconds
(without alignments)
1116.583 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 1954
Sequence: 1 ACMVATRSQTDTLEKQVERA.....KNDKSLSEKSKIQLLEAK 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	397	20.3	539	1 M24_STRPY	P12379 streptococ
2	385.5	19.7	492	1 M5_STRPY	P02977 streptococ
3	237.5	13.2	483	1 M6_STRPY	P08089 streptococ
4	212.5	10.9	1433	1 REST_CHICK	O42184 gallus gall
5	212.5	10.9	1526	1 MYS2_SCHPO	Q9usi6 schizosacch
6	206.5	10.6	2349	1 TPR_HUMAN	P12270 homo sapien
7	203	10.4	1969	1 MYSA_CAEEL	P12844 caenorhabdi
8	202	10.3	1427	1 REST_HUMAN	P30622 homo sapien
9	201.5	10.3	1102	1 MYSC_CHICK	P29616 gallus gall
10	200	10.2	1208	1 YDH6_SCHPO	Q92351 schizosacch
11	200	10.2	1939	1 MYSA_HUMAN	P13333 homo sapien
12	199	10.2	1509	1 MYSN_ACACA	P05659 acanthamoeb
13	197	10.1	1939	1 MYSA_MESAU	P13339 mesocricetu
14	195	10.0	1937	1 MYSP_HUMAN	P13335 homo sapien
15	195	10.0	1938	1 MYSA_MOUSE	Q02566 mus musculu
16	193.5	9.9	1875	1 MLPI_YEAST	Q02455 saccharomyc
17	193.5	9.9	1934	1 MYSB_MESAU	P13540 mesocricetu
18	193.5	9.9	1935	1 MYSB_PIG	P79293 sus scrofa
19	193.5	9.9	1957	1 YD86_SCHPO	Q10411 schizosacch
20	193	9.9	875	1 ZIPI_YEAST	P13111 saccharomyc
21	193	9.9	1790	1 USOI_YEAST	P25386 saccharomyc
22	192.5	9.9	1935	1 MYSB_HUMAN	P12883 homo sapien
23	192.5	9.9	1935	1 MYSB_RAT	P02564 rattus norv
24	191.5	9.8	2245	1 MYSJ_DICDI	P54697 dictyosteli
25	191.5	9.8	2663	1 CENE_HUMAN	Q02224 homo sapien
26	190	9.7	944	1 NUFI_YEAST	P03280 saccharomyc
27	189.5	9.7	866	1 MYSP_SCHJA	Q05870 schistosoma
28	189.5	9.7	2022	1 ANTI_ONCVO	P21249 onchocerca
29	189	9.7	866	1 MYSP_SCHMA	P06198 schistosoma
30	189	9.7	1928	1 MYSL_YEAST	P08964 saccharomyc
31	189	9.7	1938	1 MYSA_RAT	P02563 rattus norv
32	188.5	9.6	3210	1 CENE_HUMAN	P49454 homo sapien
33	187.5	9.6	2649	1 BPAL_HUMAN	Q03001 homo sapien

34	186.5	9.5	756	1 Y328_MYCGE	Q49419 mycoplasma
35	186.5	9.5	1972	1 MYST_RABIT	P35748 oryctolaqu
36	186.5	9.5	1976	1 MYSO_HUMAN	P35580 homo sapien
37	186	9.5	465	1 MYSA_RABIT	P04460 oryctolaqu
38	186	9.5	2116	1 MYS2_DICDI	P08799 dictyosteli
39	185	9.5	976	1 SCPI_HUMAN	Q15431 homo sapien
40	184.5	9.4	1033	1 Y328_MYCPN	P75310 mycoplasma
41	184	9.4	539	1 MY53_HYDAT	P39922 hydra atten
42	184	9.4	863	1 MYSP_TAESO	P35418 taenia soli
43	184	9.4	1940	1 MYSE_CHICK	P02565 gallus gall
44	183.5	9.4	692	1 MYS_PODCA	Q05D00 podocoryne
45	183.5	9.4	1539	1 Y373_HUMAN	O15078 homo sapien

ALIGNMENTS

RESULT 1	
M24_STRPY	
ID M24_STRPY	STANDARD; PRT; 539 AA.
AC P12379:	
DT 01-OCT-1989 (Rel. 12, Created)	
DT 01-OCT-1989 (Rel. 12, Last sequence update)	
DT 01-OCT-1996 (Rel. 34, Last annotation update)	
DE M PROTEIN, SEROTYPE 24 PRECURSOR.	
GN EM24.	
OS Streptococcus pyogenes.	
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
OC Streptococcus.	
OX NCBI_TaxID=1314;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=A24 VAUGHN;	
RX MEDLINE=88115166; PubMed=3276665;	
RA Mouw A.R., Beachey E.H., Burdett V.;	
RT "Molecular evolution of streptococcal M protein: cloning and	
RT nucleotide sequence of the type 24 M protein gene and relation to	
RT other genes of Streptococcus pyogenes.";	
RL J. Bacteriol. 170:676-684(1988).	
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES	
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF	
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO	
CC PHAGOCYTOSIS.	
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.	
CC -!- SIMILARITY: TO OTHER M PROTEINS.	
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS	
CC IN THE REGION OF THE MEMBRANE ANCHOR.	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	
CC -----	
CC EMBL; M19031; AA226874.1; -	
CC PIR; A28549; A28549.	
CC InterPro; IPR001899; -	
CC Pfam; PF00746; Gram_pos_anchor; 1.	
CC PRINTS; PR00015; GPOSANCHOR.	
CC PROSITE; PS00343; GRAM_POS_ANCHORING; 1.	
CC Signal; Virulence; Cell wall; Transmembrane; Phagocytosis;	
CC Duplication; Repeat; Antigen; Coiled coil.	
CC SIGNAL	
CC 1 42	
CC POTENTIAL.	
CC CHAIN 43 539	
CC M PROTEIN, SEROTYPE 24.	
CC DOMAIN 43 514	
CC EXTRACELLULAR (POTENTIAL).	
CC TRANSMEM 515 534	
CC MEMBRANE ANCHOR.	
CC DOMAIN 535 539	
CC CYTOPLASMIC (POTENTIAL).	
CC DOMAIN 118 301	
CC 5.3 x 35 AA TANDEM REPEATS, A-TYPE.	
CC REPEAT 118 152	
CC A-1.	
CC REPEAT 153 187	
CC A-2 (1 ALTERATION).	
CC REPEAT 188 222	
CC A-3 (2 ALTERATIONS).	


```

CC CC -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS ARE PRODUCED BY
CC CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; AF014012; AAC60344.1; -
DR DR EMBL; AF020764; AAC60345.1; -
DR DR EMBL; AF045650; AAC03547.1; -
DR DR EMBL; AF045651; AAC03548.1; -
DR DR InterPro: IPR000938; -
DR DR Pfam: PF01302; CAP_GLY; 2
DR DR PROSITE; PS00845; CAP_GLY_1; 2
KW Cytoskeleton; Microtubules; Coiled coil; Alternative splicing.
FT DOMAIN 79 121
FT DOMAIN 144 207
FT DOMAIN 235 277
FT DOMAIN 305 332
FT DOMAIN 351 1353
FT DOMAIN 1414 1427
FT VARSPLIC 458 492
FT VARSPLIC 458 492
FT VARSPLIC 803 803
FT VARSPLIC 458 458
FT CONFLICT 309 309
FT CONFLICT 440 440
FT SEQUENCE 1433 AA; 161026 MW; 5631CB8683498E23 CRC64;
SQ
Query Match 10.9%; Score 212.5; DB 1; Length 1433;
Best Local Similarity 21.4%; Pred. No. 0.0041;
Matches 92; Conservative 94; Mismatches 181; Indels 63; Gaps 13;
QY 5 ATRSDTDTLEKVRADKFEIENNTLKKSDLSFNKALKDHNDLTELSENAKELRG 64
DB 854 AQRAMQETINKNGEQFALMSEQLKSNLTVMETKLE-REEREQQLTEAKVKLEN 912
QY 65 -----SAVTRGTINDPRAKE-ALDKVELENHDLTKNEGL-----KTENE 104
DB 913 DIAEIMKSSGSSAQLMKMDELRLKLEQLQLE---LTKANEKAVQLQKNVEQTAQK 969
QY 105 GLKTENEGKLTENGLTEVDV--FPRGVENPDKARELLNLYDVENSMLQANDKLPW 162
DB 970 AEQSQQETLTKHQBELKKMQDQLTDMKKOMETSONQYKDLQAKYKETSEMITKHDADIK 1029
QY 163 RVRYTRHTPEDKL---KIIDDLDAHEHLQOQNKSLQNGDGNPREVIEDL----- 212
DB 1030 GFKNLLDAEALKAQKNDLQETAEELKQAEQ---AKADKRAEVLQTMKVTKEK 1086
QY 213 -AANNPAIQIRLHENKDLKARLENAMVAGDFKRAAGTLLDQVTLTKHNSNYQQYN 271
DB 1087 DAHQEIETLASLENSRQTNKELQNLQNLQNLKQNLKNEELTKSKELNLENKKEELK 1146
QY 272 A--QAGRLDRQAEYKLGUNDAERLLQELNIDVATRSQTDPLEKVRADKFEIENNT 329
DB 1147 KEFEALKAAQKSQLAALQAEENVKLAELG---RSRDEVTSHQLEE--ERSVLNNQL 1201
QY 330 LKLNKSLSF-----NNKALKDHNDLTELSENAKELRKNDKSLSEKA 373
DB 1202 LEMKRETLKKEIDERASLQKSI SDTSALITQKDEELEKRLNEITVLRGENASATLQ 1261
QY 374 SKIQLEARK 383
DB 1262 SVKLTLESDK 1271

```

```

RESULT 5
MYS2_SCHPO STANDARD; PRT; 1526 AA.
ID MYS2_SCHPO
AC Q9US16; P78969;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN II HEAVY CHAIN.
GN MYO2 OR SPCC645.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=972;
RX MEDLINE=98075862; PubMed=9415380;
RA May K.M., Watts F.Z., Jones N., Hyams J.S.;
RT "Type II myosin involved in cytokinesis in the fission yeast,
RL Schizosaccharomyces pombe";
RN Cell Motil. Cytoskeleton 38:385-396(1997).
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE
CC CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL; U75357; AAC49908.1; -
DR DR EMBL; AL049498; CAB39901.1; -
DR DR HSSP; P08799; 1MND.
DR DR InterPro; IPR000048; -
DR DR InterPro; IPR001609; -
DR DR PRINTS; PR00193; MYOSINHEAVY.
DR DR PROSITE; PS50096; IQ; 1.
KW Myosin; Coiled coil; Actin-binding; ATP-binding; Alkylation.
FT DOMAIN 1 815
FT DOMAIN 816 1526
FT DOMAIN 758 787
FT DOMAIN 875 1244
FT NP_BIND 170 177
FT DOMAIN 634 656
FT DOMAIN 734 748
FT MOD_RES 674 674
FT CONFLICT 1337 1337
FT SEQUENCE 1526 AA; 176430 MW; D71D51D6578192BA CRC64;
SQ
Query Match 10.9%; Score 212.5; DB 1; Length 1526;
Best Local Similarity 22.1%; Pred. No. 0.0044;
Matches 93; Conservative 74; Mismatches 161; Indels 93; Gaps 12;
QY 11 DTLEKVRADKFEIENNTLKKSDLSFNKALKDHNDLTELSENAKELRGSAVTRG 70
DB 880 ETLRRTQERLANIEDSFSETKQONLQRESASLKQINNESEL-----LEKTSKVT 933
QY 71 TINDPQRAKALDKVELENHDLTKNEGLTKENEGLKTE---NEGLKTENEGKLTVEDR 126
DB 934 LLSEONELKELSLSEKDLDTKGELESRENNATVLSKAEFNEQCKSLQETIVTKDAE 993

```

```
QY 127 VPRGTVPKARELLNKYDVNSMLQANNDKLPWRVRYTRHTPEDKLKKI-----178
Db 994 L-----DKLKYSIDYKTEIQEMRLNOKNNEKSIOQESLSLRVKKLERENS 1044
QY 179 --IDDLDAKEHELOQNEKLSLQNGDNPREVIDLAANNPAINQIRLHSHENKDLKARLE 236
Db 1045 TLISDVST-----LKOQKEELSVLKG-----VOELTINN-----LE 1075
QY 237 NAMEVAGRDFKRGCTL-----LDQVTLQYTKHNSNYQQ-----YNAQAGRDLRLQ 281
Db 1076 EKVNYLEADYKQLPKLKELESNDKDLQLOQATKNKEAKVKECLNNIKSLTKELEN 1135
QY 282 KAEYKGLNDWAERLLQELNIDVATRSQDTDLKQVQERADKFEIENNTLK-----331
Db 1136 KEERQCNLSASLKYIELQETHENLLKVSDELENKYYKKEGLQDLLEGLKVDVTNFQELS 1195
QY 332 LKNSDLSFNKAL-----KHNDDEL-----EELSNAKEKLKRNKDSLSKSKIOLEA 381
Db 1196 KHRDLTFNHESLLRSQASYSYKEXLSLASSENKDLNKNVSLTKQVNELSPKASKVPPEL 1255
QY 382 R 382
Db 1256 K 1256

RESULT 6
TPR_HUMAN
ID TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NUCLEOPROTEIN TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RN REVISIONS, AND CHARACTERIZATION.
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
RN SEQUENCE OF 1-142 FROM N.A.
RX MEDLINE=88262257; PubMed=3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RT "tpr homologues activate met and raf.";
RL Oncogene 2:617-619(1988).
CC -1- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P62.
CC -1- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
CC BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
CC -1- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
CC OR RAF GENES.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X66397; CAA47021.1; -
DR EMBL; Y00672; CAA68681.1; -
DR PIR; S00928; S00928.
DR MIM; 189940; -
KW Heptad repeat pattern; Coiled coil; Proto-oncogene;
KW Chromosomal translocation; Nuclear protein; Transport.
FT DOMAIN 78 360
FT COILED COIL (POTENTIAL).
FT DOMAIN 422 571
FT COILED COIL (POTENTIAL).
FT DOMAIN 575 628
FT COILED COIL (POTENTIAL).
FT DOMAIN 758 805
FT COILED COIL (POTENTIAL).
FT DOMAIN 834 869
FT COILED COIL (POTENTIAL).
FT DOMAIN 934 979
FT COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064
FT COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166
FT COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241
FT COILED COIL (POTENTIAL).
FT DOMAIN 1262 1304
FT COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434
FT COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595
FT COILED COIL (POTENTIAL).
FT DOMAIN 527 530
FT POLY-SER.
FT DOMAIN 1833 1836
FT POLY-GLU.
FT DOMAIN 1957 1964
FT POLY-ASP.
FT DOMAIN 2295 2298
FT POLY-SER.
SQ SEQUENCE 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;
```

```
Query Match 10.6%; Score 206.5; DB 1; Length 2349;
Best Local Similarity 21.4%; Pred. No. 0.013;
Matches 102; Conservative 80; Mismatches 174; Indels 121; Gaps 17;
QY 5 AFRSDTDTLEKQVER----ADKFEIENNTLKNSDLSFNKALKDHDELTELSNAKE 60
Db 91 ALTEKKELETAQDNIAIQSFRTKEELEAKRDLRTNRELSELEYEDVKRINE 150
QY 61 KLKRSVTRGTINDPQRAKEALDYELNHLKTKNEGLKTENEGKLTENEGKLTENEG 120
Db 151 KLKESNTTKGEL-----QLKLDLQASDVSVKVRKLEQEKELHSHQNTWLATE---L 201
QY 121 KTEVDVRVPRG-----TVENP-----DKARELLNKYDVNSMLQANNDKLPWRV 164
Db 202 KTKTDELLALGREGNEILELKNLENKKEVSRLEEQMNGLKTNSNEHLQKHVEDLLTKL 261
QY 165 RYTRH---TPEDKL-----KKTIIDDLDAKEHELOQNEKLSLQNGDNPRE 207
Db 262 KEAKEQQASMEKPFHNLNAHLKSLNLYKSAADSEAKSNELTRAVEL-----HK 312
QY 208 VIEDLAANNPAINQIRLR-HENKD-----LKALENAMVAGRDFKRGATLLD 254
Db 313 LLKEAGEANKAIQDHLEVEQSKDQMEKLEKIGRLEKELENANDLLSAT-KRKGAIS 371
QY 255 Q-----VTOLYTKH-----NSNYQQYNAQAGRDL 279
Db 372 EELAAASPTAAAVAKIVPKMKTLEYNAVFQDQLLEKLENKINKLYDVIK-EV 430
QY 280 ROKAEYKGLNDWAERLLQEL-NIDVATRSQDTDLKQVQERADKFEIENNTLKNSDLS 338
Db 431 EKAPILKQRQEEYERAKAVASLSVKLEQAMKEIQRLQEDTDKANKQSSVLERNRME 490
QY 339 FNNKALDHNDDELFEELSNAK-----EKLKNDKSLSEKA-----SKIOELE 380
Db 491 IOVKDLSQIIRVLLMELEEARGNHVIRDEEVSSADISSSEVSIQHLVSYRNIBELQ 547

RESULT 7
MYSACAEEL
ID MYSACAEEL STANDARD; PRT; 1969 AA.
AC P12844;
```

01-OCT-1989 (Rel. 12, Created)
 01-OCT-1989 (Rel. 12, Last sequence update)
 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MYOSIN HEAVY CHAIN A (MHC A).
 GN MYO-3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=89178677; PubMed=2926820;
 RA Dibb N.J., Maruyama I.N., Krause M., Karn J.:
 RT "Sequence analysis of the complete Caenorhabditis elegans myosin
 heavy chain gene family.";
 RL J. Mol. Biol. 205:603-613(1989).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PPM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMN) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C.
 ELEGANS.
 CC -1- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY
 WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
 CC -1- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
 SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
 HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
 THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
 THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: X08067; CAA30856.1; -
 PIR: S02771; S02771.
 DR HSSP: P08799; LMND.
 DR InterPro: IPR000048; -
 DR InterPro: IPR001609; -
 DR InterPro: IPR002928; -
 DR Pfam: PF00612; IQ; 1.
 DR Pfam: PF01576; Myosin_tail; 1.
 DR Pfam: PF00063; myosin_head; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
 KW Multigene family.
 FT DOMAIN 1 851 GLOBULAR HEAD (S1).
 FT DOMAIN 852 1969 RODLIKE TAIL (S2 AND LMN DOMAINS).
 FT DOMAIN 857 1969 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP.
 FT DOMAIN 667 689 ACTIN-BINDING.
 FT DOMAIN 770 784 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 707 707 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 717 717 ALKYLATION (SH-2) (POTENTIAL).
 SEQUENCE 1969 AA; 225509 MW; 64577BBAF7EAD80A CRC64;

Query Match 10.4%; Score 203; DB 1; Length 1969;
 Best Local Similarity 19.2%; Pred. No. 0.016;
 Matches 101; Conservative 96; Mismatches 165; Indels 164; Gaps 18;
 QY 1 ACWATRSQTDLTLEKVRQADKFEIENNTLKLKNSDLSFNKALKDHNDLT-----E 53
 DB 899 ALFSLSETEKANLADAEERNEKLNOLKATLESKLSDIINGQLEDQERNEDLARQKKTDQ 958
 QY 54 ELSNAKEKLRSVAVTRGTINDPQRAKEALDKYELENHDLKTKYNEGLT----- 106
 DB 959 ELSDTKKHVQDLEL---SLRKAQEQKSRD-----HNIRSLQDEMANQDAVAKLNKEK 1009
 QY 107 KTEENGLKTENEGKTEVDV-----PPRGTVENPDK--- 138
 DB 1010 KQEESENRLKLNEDLOSEEDKVNHLEKIRKLEQQQWDELEENIDREKSRGDIETAKRKVE 1069
 QY 139 -----ARELLN---KYDVENSIMQANNKLPWRVRYTRHTPD-----KLKKII 179
 DB 1070 GDLKVAQENIDETIKQKHVDVETTLARKEE-----LHHTNAKLAENNSIAKLQRLI 1121
 QY 180 DDLDAK-----EHELOQNEKLSLQNG-----DGNP--- 205
 DB 1122 KELTSARNAELEEAEARNRSROKSDRSRSEARELEELTERLEQQGGATAAQAANKKRE 1181
 QY 206 -----REVIEDLAANPAIONIRHEN-----KDLKARLENAMVAGRDF 246
 DB 1182 AEIAKLRRKEEDSLNHETATSSLRKRHGDVSVALETQLETQLKLUKAKSEAKSKLQKDL 1241
 QY 247 -----KRGATLLDQVTLQYTK---HNSNYQQYNAQAGRLDLRQAEY 285
 DB 1242 EESQATDSEVRSRDLKALKATIEVOYSELQTKADEOSRQLODFAALKNRLN-NENSDL 1300
 QY 286 LKGLNDWAERLLQELNIDVATRSQDTEKVOERADKFEIENNTLKLKNSDLSFNKALK 345
 DB 1301 NRSLEMDNQLNSHLKSLTQSQ---LDETRNRYDEESRERQALAAATAKLNHEHTILR 1357
 QY 346 DINDELTELSNAKEKLRENKSL-----SEKASKIQELEAR 383
 DB 1358 EHLDEAESKADLTQIKSLNAEIQWKAQFDSGLNKLLEIEAAK 1403
 RESULT 8
 REST_HUMAN
 ID REST_HUMAN STANDARD; PRT; 1427 AA.
 AC P30822;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE RESTIN (CYTOPLASMIC LINKER PROTEIN-170 ALPHA-2) (CLIP-170) (REED-
 DE STERNBERG INTERMEDIATE FILAMENT ASSOCIATED PROTEIN).
 GN RSN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Peripheral blood monocytes;
 RX MEDLINE=92289675; PubMed=1600942;
 RA Bilbe G., Delabie J., Brueggen J., Richener H., Asselbergs F.A.M.,
 RA Cerletti N., Sorg C., Odink K., Tarsay L., Wiesendanger W.,
 RA de Wolf-Peters C., Shipman R.;
 RT "Restin: a novel intermediate filament-associated protein highly
 expressed in the Reed-Sternberg cells of Hodgkin's disease.";
 RL EMBO J. 11:2103-2113(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92405160; PubMed=1356075;
 RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
 RT "CLIP-170 links endocytic vesicles to microtubules.";
 RL Cell 70:887-900(1992).

CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
CC CYTOSKELETON.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64838; CAA46050.1; -;
CC EMBL; M97501; AAA35693.1; -;
CC PIR; S22695; S22695.
CC MIM; 179838; -;
CC InterPro; IPR000938; -;
CC Pfam; PF01302; CAP_GLY; 2.
CC PROSITE; PS00845; CAP_GLY_1; 2.
CC Cytoskeleton; Microtubules; Coiled coil; Alternative splicing.
CC
CC DOMAIN 60 125
CC FT DOMAIN 143 204
CC FT DOMAIN 214 279
CC FT DOMAIN 304 331
CC FT DOMAIN 350 1342
CC FT DOMAIN 1408 1421
CC FT VARSPLIC 457 491
CC FT CONFLICT 1069 1069 D -> E (IN REF. 2).
CC SQ SEQUENCE 1427 AA; 160989 MW; 0A4F16DD94254E8 CRC64;

Query Match 10.3%; Score 202; DB 1; Length 1427;
Best Local Similarity 21.4%; Pred. No. 0.012;
Matches 95; Conservative 86; Mismatches 161; Indels 102; Gaps 18;

Qy 7 RSQDTLLEKQVRAKDFEINNTLK-----LNKSDLSFNKALKHNDLTELSENAKEK 61
Db 861 RSMQETVNLKHQEEQFMLSDDLEKLENLADWEAKFREK-----DEREQLIKAKEK 914
Qy 62 L-----RGSAYTRGTINDPQRAKEALDKYLENHDLKTKNEGLKTENGLKT 108
Db 915 LENDIAEIMKMSGDNSQLTK--MNDELRLKER-----DVEELQLK-----LTK 956
Qy 109 ENEG---LKTENGLKTEVDVPRGTVPNDKARELLNKY-DVENSMLQANDKLPWRV 164
Db 957 ANENASFLOKSIEDMTVKAESQQAQKHEEKELERKLSLEKMKETSHQCOQLKA 1016
Qy 165 RYRHTRPEDKIL--KIIDDL-----AKE-----HELQOONKELSLQNG 201
Db 1017 RYERATSETKTHEEILQNLQKTLTDTEKLGAREENSGLLQLEELRLQADKAAQAQ 1076
Qy 202 DGNPREVIEDLAANPAQNTLRHSHENKOLKARLENAMVAGRDFKRACTLLDQVTLTY 261
Db 1077 AEDAMQIMEQTKK--TETLASIEDTKQTNALQNELDTLKENNKLKNEELNKRSELL 1134
Qy 262 KHNSNYQQYNAQAQGLD--LRQAEYLKGLNDWAERLLQELINIDVATRSQDTLLEKQV 319
Db 1135 VENQKMEFREIEITLQKAAQKSQLQAEENVKLAELG---RSRDEVTSHQKLEE- 1190
Qy 320 ADKFEIENNTLKNLSDFNKKALKDHND-----LT-----EELSNK 359
Db 1191 -ERSVNLNOLLEMKRE-----SKFIDADEKASLOKSISITSALLTEKDAEKLKRE 1245
Qy 360 EKLKNDKSLSEKASKIQELEARK 383
Db 1246 TVLRGENASAKSLHVVQTLSEDK 1269

RESULT 9
ID MYSC_CHICK STANDARD; PRT; 1102 AA.
AC P29616;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ISOFORM (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 65-1102 FROM N.A.
RX STRAIN=Broiler breeders; Tissue=Heart;
RX MEDLINE=92130260; PubMed=1774788;
RA Stewart A.F.R., Camoretti-Mercado B., Perlman D., Gupta M.,
RA Jakovcic S., Zak R.;
RT "Structural and phylogenetic analysis of the chicken ventricular
RT myosin heavy chain rod".
RL J. Mol. Evol. 33:357-366(1991).
RN [2]
RP SEQUENCE OF 1-259.
RX Tissue=Heart;
RX MEDLINE=93039740; PubMed=1418675;
RA Watanabe B.;
RT "Amino-acid sequence of the short subfragment-2 in adult chicken
RT cardiac muscle myosin".
RL Biol. Chem. Hoppe-Seyler 373:1045-1054(1992).
CC -!- FUNCTION: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- TISSUE SPECIFICITY: VENTRICULAR MUSCLE AND TRACES LEVELS IN THE
CC ATRIUM. ALSO TRANSIENT EXPRESSION IN SKELETAL MUSCLE DURING FETAL
CC DEVELOPMENT AND REGENERATION FOLLOWING FREEZE INJURY.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
CC CONSERVED.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59552; CAA42130.1; -;
CC HSP; P03069; IZIK.
CC InterPro; IPR002928; -;
CC Pfam; PF01576; Myosin_tail; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Heptad repeat pattern; Multigene family.
CC
CC NON_TER 1
CC DOMAIN <1 1102 RODLIKE TAIL (S2 AND LMM DOMAINS).
CC DOMAIN <1 1102 COILED COIL (POTENTIAL).
CC CONFLICT 65 65 A -> T (IN REF. 1).
CC CONFLICT 92 92 V -> A (IN REF. 1).
CC CONFLICT 113 114 CS -> FAL (IN REF. 1).
CC CONFLICT 135 135 H -> Q (IN REF. 1).
CC CONFLICT 185 185 T -> R (IN REF. 1).
CC SEQUENCE 1102 AA; 128008 MW; 2293668D049825DC CRC64;

```

SQ SEQUENCE 1208 AA; 140763 MW; 70264159ADD42424 CRC64;

Query Match 10.2%; Score 200.; DB 1.; Length 1208;
Best Local Similarity 21.6%; Pred. No. 0.013;
Matches 89; Conservative 77; Mismatches 160; Indels 86; Gaps

QY 3 MVATRSQDTLEKVOERADKFEIENNTLKLNKSDLSFNKALKDHNDLDELTEELSNAK--- 59
   | | : : : : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 427 MLQPSLEERDVLSEKLOTLDDNNSLRMTSSLGNOIESLRQREIDEKNHLRLLA 486
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 60 EKLRSVTRCTINDPORAEALDKYLENHDLKTNEGGLKTNEGGLKTNEGGLKTNEG 119
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 487 SKNSDKALATNRKLQEVTK-----ELET--LRMKNSDNLNIHDLREENEGLTKIDS 538
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 120 LKTEVDFVPRGTVPNDKARELLNKVDVENSMLQANNDKLPWRVYRTHPTDCLKKII 179
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 539 ITRKEDRLI-----NELEQRKISYEVNVSELNGTIDE---YRNKLKDRETYNEVM 586
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 180 DDLDAKHEHLOQONEKLSLQNGDGNPREVIEDLAANPAIQNLRHKNKDKLARLENAM 239
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 587 NAFQYKQNDLRRFHESIN-----KLODREKELTSNLEKKN 621
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 240 EVAGRDPKFRAGTLLDQVLTQYTKHNSYQYNAQAGRLDLRQKAEVLLKGLNDWAERLLQ 299
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 622 LV-----ISSLRVAMLEKERESIKKYLGNADQLD---NTNMEILNDKISVLQKQ 671
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 300 L-----NIDVATRSQDTTLEKVOERADKFEI-----ENNLTALKNSDLS---FNNKALKDHN 348
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 672 LTVDKDELQVSEEREAEIVAGOKLSASFELMSNEKQALEKLYSSLKNELINAQNLDDR 731
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 349 DELTEEL-----SNAKEKLRKN-----DKSLSEKASKIOELEARK 383
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 732 EEELSLSKKLFERKIRSGNSDDIEKNEINVLSNDELAKLQAIHLES DK 783

RESULT 11
MYSA_HUMAN
ID MYSA_HUMAN STANDARD; PRT; 1939 AA.
AC P13533; Q13943; Q14906; Q14907;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE ALPHA ISOFORM.
GN MYH6 OR MYHCA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=921133665; PubMed=1776652;
RY Matsuoka R., Beisel K.W., Furutani M., Arai S., Takao A.;
RT "Complete sequence of human cardiac alpha-myosin heavy chain gene and
RT amino acid comparison to other myosins based on structural and
RT functional differences.";
RL Am. J. Med. Genet. 41:537-547(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94140346; PubMed=8307559;
RY Epp T.A., Dixon I.M., Wang H.Y., Sole M.J., Liew C.-C.;
RT "Structural organization of the human cardiac alpha-myosin heavy
RT chain gene (MYH6).";
RL Genomics 18:505-509(1993).
RN [3]
RP SEQUENCE OF 1-177 AND 1551-1939 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RY Yamauchi-Takihara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RY Yamauchi-Takihara K., Sole M.J., Liew J., Ing D., Liew C.-C.;

```

Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).

[5] SEQUENCE OF 1407-1939 FROM N.A.

RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;

RT *Molecular cloning and characterization of human cardiac alpha- and

RT beta-form myosin heavy chain complementary DNA clones. Regulation of

RT expression during development and pressure overload in human

RT atrium*;

RL J. Clin. Invest. 82:524-531(1988).

CC -!- FUNCTION: MUSCLE CONTRACTION.

CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.

CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY

CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT

CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE

CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED

CC SUBFRAGMENT (S2).

CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE

CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.

CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,

CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE

CC CONSERVED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL; D00943; BAA00791.1; -

DR EMBL; Z20636; CAA79675.1; -

DR EMBL; M25140; AAA60386.1; -

DR EMBL; M25162; AAA60386.1; JOINED.

DR EMBL; M25142; AAA60387.1; -

DR EMBL; M25141; AAA60387.1; JOINED.

DR EMBL; M21664; AAA36344.1; -

DR PIR; A46762; A46762.

DR PIR; A28908; A28908.

DR HSP; P08799; LMND.

DR MIM; 160710; -

DR InterPro; IPR000048; -

DR InterPro; IPR001609; -

DR InterPro; IPR002928; -

DR Pfam; PF00612; IQ; 1.

DR Pfam; PF01576; Myosin_tail; 1.

DR Pfam; PF00063; myosin_head; 1.

DR PRINTS; PR00193; MYOSINHEAV.

DR PROSITE; PS50096; IQ; 1.

DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;

DR ATP-binding; Methylation; Alkylation; Heptad repeat pattern;

DR Multigene family.

DR DOMAIN 1 841 GLOBULAR HEAD (S1).

DR DOMAIN 842 1939 RODLIKE TAIL (S2 AND LMM DOMAINS).

DR DOMAIN 842 1939 COILED COIL (POTENTIAL).

DR NP_BIND 178 185 ATP.

DR DOMAIN 657 679 ACTIN-BINDING.

DR DOMAIN 759 773 ACTIN-BINDING.

DR MOD_RES 129 129 METHYLATION (SH-1) (POTENTIAL).

DR MOD_RES 697 697 ALKYLATION (SH-2) (POTENTIAL).

DR MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).

DR CONFLICT 88 88 Q -> E (IN REF. 1).

DR CONFLICT 574 574 Q -> P (IN REF. 1).

DR CONFLICT 608 608 A -> G (IN REF. 1).

DR CONFLICT 744 744 T -> A (IN REF. 1).

FT CONFLICT 790 790 M -> I (IN REF. 1).

FT CONFLICT 1014 1014 V -> A (IN REF. 1).

FT CONFLICT 1021 1021 S -> T (IN REF. 1).

FT CONFLICT 1101 1101 A -> V (IN REF. 1).

FT CONFLICT 1290 1290 A -> S (IN REF. 1).

FT CONFLICT 1373 1373 W -> C (IN REF. 1).

FT CONFLICT 1533 1533 K -> N (IN REF. 5).

FT CONFLICT 1540 1540 L -> M (IN REF. 5).

FT CONFLICT 1577 1578 KL -> NV (IN REF. 5).

FT CONFLICT 1705 1706 EQ -> DR (IN REF. 1).

FT CONFLICT 1733 1733 E -> D (IN REF. 1).

FT CONFLICT 1734 1734 A -> S (IN REF. 2).

FT CONFLICT 1737 1737 T -> S (IN REF. 1).

FT CONFLICT 1763 1763 D -> H (IN REF. 1).

FT CONFLICT 1788 1788 M -> I (IN REF. 3).

FT CONFLICT 1871 1871 D -> N (IN REF. 5).

FT CONFLICT 1882 1882 R -> G (IN REF. 5).

FT CONFLICT 1890 1890 Q -> R (IN REF. 5).

FT CONFLICT 1933 1933 MISSING (IN REF. 5).

SQ SEQUENCE 1939 AA; 223689 MW; ECB87E7CE8768B6F CRC64;

Query Match 10.2%; Score 200; DB 1; Length 1939;

Best Local Similarity 22.2%; Pred. No. 0.022;

Matches 98; Conservative 69; Mismatches 141; Indels 134; Gaps 18;

QY 7 RSOTDTLEKQERADKFEIENNTLKNSDLSFNKALKDHNDLTELSELSNAEKLGRS- 65

DB 894 QAEQNDLNDRECDQLKKNKIQLEAKVEM---NERLED-EEEMNAELTAKRKLEDEC 949

QY 66 AVTRGTINDPQRAEKALDKYLENHDLTKNGLKTENEGI-----YTENEGLK 114

DB 950 SELKKDIDD---LELTAKVKEKEKHATENKVNLTTEEMAGLDEITIAKLTREKALQEAHQ 1006

QY 115 TENEGLTQVDRVPP-----RGTVENPDKARELLNKYD-----VEN 150

DB 1007 QALDDQLQVEDKVNLSKSKVKVLEQQVDDLEGSLEQKQKVMDLERAKRKLEGLDKLTQE 1066

QY 151 SMLQANNDKLPWRVRYTRHTPEDKLLKIIDDLDAKEHELOQONEKL-----SL 198

DB 1067 SIMDLENDKQL-----EEKLKK-----KEFDINQNSKIEDEQALALQLOKKL 1110

QY 199 QNGDGNPREVIEDLAANPA---IQNTR--LRHENKDLKARLENAMAVEAGDFKRACTLL 253

DB 1111 KENQARIEELEEELEAERTARAKVEKLSRSLSELESELEEA-----GGATS 1160

QY 254 DQVTOLYTKHNSNYQYNAQAGRLDLROKAYLKLNDWAERLLQELNIDVATR----- 307

DB 1161 VOI-----EMNKKREAEFOKMRDRLEATLQHEATAALRKKHADS 1201

QY 308 -----SOTDTLEKQERADKFEIENNTLKNSDLSFNKALKDHNDLTELSELSNAEKL 362

DB 1202 VAELEGEQIDNLRQVKQKLEK---EKSEFKLEDDVTSN-----MQIITAKANL 1247

QY 363 RKNDKSLSEKAS-----KIQELE 380

DB 1248 EKVSRTLEQANRYRKLEEAQ 1269

RESULT 12

MYSN_ACACA

ID MYSN_ACACA STANDARD; PRT; 1509 AA.

AC P05659;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE MYOSIN II HEAVY CHAIN, NON MUSCLE.

OS Acanthamoeba castellanii (Acanthamoeba).

OC Eukaryota; Acanthamoebidae; Acanthamoeba.

OX NCBI_TaxID=5755;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87308395; PubMed=3040773;

CC CONSERVED.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M76598; AAA37159.1; -
CC EMBL: M76599; AAA37160.1; -
CC EMBL: M76600; AAA37161.1; -
CC EMBL: M76601; AAA37162.1; -
CC EMBL: M62404; AAA37424.1; -
CC HSSP: P08799; 1MND.
CC SWISS-2DPAGE: Q02566; MOUSE.
CC MGD; MGI:97255; Myhca.
CC InterPro: IPR000048; -
CC InterPro: IPR001609; -
CC InterPro: IPR002928; -
CC Pfam: PF00612; IQ; 1.
CC Pfam: PF01576; Myosin_tail; 1.
CC Pfam: PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC PROSITE; PS50096; IQ; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
CC Multigene family; Calmodulin-binding; Polymorphism.
CC DOMAIN 1 841 GLOBULAR HEAD (S1).
CC FT DOMAIN 842 1938 RODLIKE TAIL (S2 AND LWM DOMAINS).
CC FT DOMAIN 842 1938 COILED COIL (POTENTIAL).
CC FT NP_BIND 178 185 ATP.
CC FT DOMAIN 657 679 ACTIN-BINDING.
CC FT DOMAIN 759 773 ACTIN-BINDING.
CC FT DOMAIN 790 807 CALMODULIN-BINDING (BY SIMILARITY).
CC FT DOMAIN 816 833 CALMODULIN-BINDING (BY SIMILARITY).
CC FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
CC FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
CC FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
CC FT VARIANT 194 194 Y -> D.
CC FT VARIANT 545 545 Y -> A.
CC FT VARIANT 838 838 I -> S.
CC SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;

Query Match 10.0%; Score 195; DB 1; Length 1938;
Best Local Similarity 21.9%; Pred. No. 0.037;
Matches 97; Conservative 70; Mismatches 141; Indels 134; Gaps 18;

Qy 7 RSQDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELTELSNAKEKLRS- 65
Db 894 QAEQDNDAEERCQDLIKKKIQLEAKVKEMT---EKLED-EEMNAELTAKRKLEDEC 949
Qy 66 AVTRGTINDPQRAKEALDKYLENHDLTKEGKLTENEGL-----KTENEGLK 114
Db 950 SELAKDIDD---LELTAKVEKEKHATENVKVNLTEEMAGLDEIIAKLTKEKKALQEAHQ 1006
Qy 115 TENEGKTEVDVRVP-----RGIVENPDKARELLNKYD-----VEN 150
Db 1007 QALDDLOAEEDKVNTLTAKSKVLEQQVDDLEGSLEQEKVKYRMDLERAKRKLGLDKLTQE 1066
Qy 151 SMLQANNDKLPWRYRTHPTEDKLLKIIDDLDAKEHELOQQNEKL-----SL 198
Db 1067 SIMDLENDKLOL-----EELKK-----KEFDISQQNSKIEDEQALALQLOKKL 1110
Qy 199 QNGDGNPREVIEDLAANPA---TONTR--LRHENKDLKARLENAMVAGRDFKRACTLL 253
Db 1111 KENQARIEELEEELEAERTAKRVEKRLRSLSRELEETSERLEEA-----GGATS 1160
Qy 254 DQVTOLYTKHNSYQOYNAQAGRLDLRQAEYKLGNDWAERLLQELNIDVATR----- 307
Db 1161 VOI-----EMNKKRAEFOKMRDLEATLOHEATAAALRRKHADS 1201

Qy 308 -----SQTDTLEKVOERADKFEIENNTLKLKNSDLSFNKALKDHNDLTELTELSNAKEKL 362
Db 1202 VAEIGEQIDNLRVKQKLER---EKSEFKLELDDVTSN-----MEQIIKAKANL 1247
Qy 363 RKNDKSLSEKAS-----KIOELE 380
Db 1248 EKVSRTLEDOANEYRVKLEEAQ 1269

Search completed: May 10, 2001, 08:40:26
Job time: 154 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 10, 2001, 08:37:38 ; Search time 27.19 Seconds
(without alignments)
1650.998 Million cell updates/sec

Title: US-09-151-409-16
Perfect score: 1954
Sequence: 1 ACWATRSQDTLEKVOERA.....KNDKSLSEKASKIQELEARK 383

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_15.*
- 1: sp.archaea.*
 - 2: sp.bacteria.*
 - 3: sp.fungi.*
 - 4: sp.human.*
 - 5: sp.invertebrate.*
 - 6: sp.mammal.*
 - 7: sp.mhc.*
 - 8: sp.organelle.*
 - 9: sp.phage.*
 - 10: sp.plant.*
 - 11: sp.rodent.*
 - 12: sp.unclassified.*
 - 13: sp.vertebrate.*
 - 14: sp.virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	966	49.4	247	2 Q53344	Q53344 streptococc
2	389	19.9	457	2 Q54510	Q54510 streptococc
3	330	16.9	272	2 Q9KGR9	Q9KGR9 streptococc
4	330	16.9	539	2 Q54719	Q54719 streptococc
5	321	16.4	355	2 Q10371	Q10371 streptococc
6	320	16.4	581	2 Q54835	Q54835 streptococc
7	317.5	16.2	231	2 Q9KGR8	Q9KGR8 streptococc
8	316	16.2	488	2 Q54830	Q54830 streptococc
9	311	15.9	484	2 Q10372	Q10372 streptococc
10	266.5	13.6	100	2 Q54632	Q54632 streptococc
11	261.5	13.4	100	2 Q54635	Q54635 streptococc
12	260.5	13.3	100	2 Q54638	Q54638 streptococc
13	259.5	13.3	100	2 Q54637	Q54637 streptococc
14	257	13.2	76	2 Q53536	Q53536 streptococc
15	257	13.2	79	2 Q54543	Q54543 streptococc
16	252	12.9	84	2 Q54598	Q54598 streptococc
17	248	12.7	484	2 Q05464	Q05464 streptococc
18	246	12.6	558	2 Q54718	Q54718 streptococc
19	237	12.1	2167	10 Q9S8U1	Q9S8U1 chara coral

ALIGNMENTS

RESULT 1	20	237	12.1	2182	10	Q9LW97	Q9LW97 chara coral
Q53344	21	229	11.7	2138	5	Q9XZE3	Q9XZE3 amoeba prot
ID Q53344	22	228	11.7	83	2	Q9ZEH6	Q9ZEH6 streptococc
AC Q53344;	23	227	11.6	2139	5	Q07569	Q07569 entamoeba h
DT 01-NOV-1996 (TReMBLrel. 01, Created)	24	226.5	11.6	1743	5	Q9G063	Q9G063 dugesia jap
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)	25	220.5	11.3	327	2	Q9ZNO8	Q9ZNO8 helicobacte
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)	26	214.5	11.0	150	2	Q54587	Q54587 streptococc
DE TETRAVALENT M PROTEIN (FRAGMENT).	27	213	10.9	1945	6	Q97757	Q97757 felis silve
GN EMM.	28	212.5	10.9	1526	3	Q9USI6	Q9USI6 schizosacch
OS Streptococcus sp.	29	212	10.8	1456	5	Q9W587	Q9W587 drosophila
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	30	208.5	10.7	443	2	Q54703	Q54703 streptococc
OC Streptococcus.	31	207.5	10.6	709	4	Q14997	Q14997 homo sapien
OX NCBI_TaxID=1306;	32	207.5	10.6	742	4	Q9UI01	Q9UI01 homo sapien
RN [1]	33	207.5	10.6	980	5	Q96246	Q96246 plasmodium
RP SEQUENCE FROM N.A.	34	207	10.6	1950	5	Q26080	Q26080 placopecten
RX MEDLINE=93346765; PubMed=8345202;	35	206.5	10.6	726	4	Q15624	Q15624 homo sapien
RA Dale J.B., Chiang E.Y., Lederer J.W.;	36	206.5	10.6	1931	13	Q91973	Q91973 coturnix co
RT "Recombinant tetraivalent group A streptococcal M protein vaccine.";	37	206.5	10.6	2363	4	Q99968	Q99968 homo sapien
RL J. Immunol. 151:2188-2194(1993).	38	205.5	10.5	1937	13	Q91BD4	Q91BD4 gallus gall
DR EMBL; S64396; AAB27745.1; -	39	204.5	10.5	911	3	Q06704	Q06704 saccharomyc
FT NON_TER 247	40	204.5	10.5	1927	5	Q25142	Q25142 halocynthia
SQ SEQUENCE 247 AA; 28252 MW; 74A5827DE945529B CRC64;	41	204	10.4	6677	5	Q9N435	Q9N435 caenorhabdi
	42	203	10.4	1827	5	Q20042	Q20042 caenorhabdi
	43	203	10.4	1946	5	Q97291	Q97291 plasmodium
	44	203	10.4	1992	5	Q21440	Q21440 caenorhabdi
	45	202.5	10.4	100	2	Q54633	Q54633 streptococc

Query Match 49.4%; Score 966; DB 2; Length 247;
Best Local Similarity 78.9%; Pred. No. 7.3e-39;
Matches 195; Conservative 0; Mismatches 0; Indels 52; Gaps 1;

```

Db 181 GTVPNDPKARELLNKYDVNSMLQANDKLPWRVRYTRHTPEDKLKTIIDLDKAHEHQ 240
QY 191 QONEKLS 197
| | | | |
Db 241 QONEKLS 247

RESULT 2
Q54510
ID Q54510 PRELIMINARY; PRT; 457 AA.
AC Q54510;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE M5.8193 PROTEIN (FRAGMENT).
GN EM5.8193.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC8193;
RX MEDLINE=94224154; PubMed=8170398;
RA Whatmore A.M., Kehoe M.A.;
RT "Horizontal gene transfer in the evolution of group A streptococcal
emul-like genes: gene mosaics and variation in Vir regulons.";
RL Mol. Microbiol. 11:363-374(1994).
DR EMBL; U02480; AAA50854.1; -.
DR INTERPRO; IPR001899; -.
DR PFAM; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 457 AA; 51172 MW; DEBBAF550AEB6379 CRC64;

Query Match 19.9%; Score 389; DB 2; Length 457;
Best Local Similarity 34.4%; Pred. No. 1.7e-11;
Matches 123; Conservative 53; Mismatches 114; Indels 68; Gaps 14;

QY 52 TEELNAKEKLRSVAVRTGINDPQRAKALDKYLELNDKLTKEGLKTEGLKTE 111
| | | | |
Db 13 TNEVS-----ATVTRGINDPQRAKALDKYLELNDKLTKEGLKTEGLKTE 64

QY 112 GLKTENGLKTEVDVPRFTGVENPNPKARELLNKYDVNSMLQANDKLPWRVRYTRHTP 171
| | | | |
Db 65 GLKTENGLKTEVDVPRFTGVENPNPKARELLNKYDVNSMLQANDKLPWRVRYTRHTP 171

QY 172 EDLKLKTIIDLDKAHEHQNEKLSLQNGD-----GNPREVIEDLAANPA 218
| | | | |
Db 112 KETLEREVQN-----TQYNNETLTKNGDLTKELNKTQELANKQESKENEKALNEL 164

QY 219 IQ-----NIRLRHENDKLARENAMVAGRD-----FKRAGTLLDQVTLTKHNS 265
| | | | |
Db 165 LEKTVKDKIAKEQENKETITGLKILDETVDKLAKEQSKQNGIKALKEALAK---KDEA 221

QY 266 NYOQYNAQAG-RDLRQKAYLKLNDWAEKRLQELNIDVATRS---QDTLEKVKQERADK 322
| | | | |
Db 222 NKISDASRKLGRDLRLDASREAKKQLEAEHQKLEQNKISEASRKLGRD-LDASREAKKQ 280

QY 323 FEIENNTLKLKNSDLSFNNAKLDHDELTEELSNKAEKLRKNDKLSKASKIQELE 380
| | | | |
Db 281 LEAEHQKLEQNKISEASRKLGR-----RDLASREAKKQVEKALEANSKLALE 331

RESULT 3
Q9KGR9
ID Q9KGR9 PRELIMINARY; PRT; 272 AA.
AC Q9KGR9;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

```

```

DE M PROTEIN (FRAGMENT).
GN EM5.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3148-00;
RT Beall B.W.;
RT "5' emm sequence emm3.2.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF288752; AAF91496.1; -.
FT NON_TER 1
FT NON_TER 272
SQ SEQUENCE 272 AA; 30862 MW; 026FB243D345A35A CRC64;

```

```

Query Match 16.9%; Score 330; DB 2; Length 272;
Best Local Similarity 46.0%; Pred. No. 5.8e-09;
Matches 81; Conservative 20; Mismatches 35; Indels 40; Gaps 6;

QY 234 RLENAMEVAGRDFFKAGTLLDQVTLTKHNSNYQYNAQAGRLDLRQKAEYLKGLNDWA 293
| | | | |
Db 41 KLKNEIE-----NLLDQVTLTKHNSNYQYNAQAGRLDLRQKAEYLKGLNDWA 90

QY 294 ERLLOELNIDVATR-----SOTDTLEK-----VQERADKFEIENNTLKLKNSDLSFNKA 343
| | | | |
Db 91 ERLLOELNGEDVKVILGVAFKDDLEKEVKELKIDKKEKYEQDLD-KDFDLAKQGVV 149

QY 344 LKD--HNDEL-----TEELSNKAEKLRKNDKLSKASKIQELE 380
| | | | |
Db 150 LSDKRHQLEKEKVKVTEATKVGQISELETVTQKVQESTMQDLTEKONRVSQL 205

```

```

RESULT 4
Q54719
ID Q54719 PRELIMINARY; PRT; 539 AA.
AC Q54719;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE M3 PROTEIN.
GN EM3.1.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE 3;
RA Dale J.B., Simmons M., Chiang E., Chiang E.;
RL Vaccine 0:0-0(0); -.
DR EMBL; U40231; AAA96960.1; -.
DR HSSP; P03069; IGCL.
DR INTERPRO; IPR001899; -.
DR PFAM; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 539 AA; 60163 MW; 93D20A4302C30F51 CRC64;

```

```

Query Match 16.9%; Score 330; DB 2; Length 539;
Best Local Similarity 46.0%; Pred. No. 1.2e-08;
Matches 81; Conservative 20; Mismatches 35; Indels 40; Gaps 6;

QY 234 RLENAMEVAGRDFFKAGTLLDQVTLTKHNSNYQYNAQAGRLDLRQKAEYLKGLNDWA 293
| | | | |
Db 55 KLKNEIE-----NLLDQVTLTKHNSNYQYNAQAGRLDLRQKAEYLKGLNDWA 104

QY 294 ERLLOELNIDVATR-----SOTDTLEK-----VQERADKFEIENNTLKLKNSDLSFNKA 343
| | | | |
Db 105 ERLLOELNGEDVKVILGVAFKDDLEKEVKELKIDKKEKYEQDLD-KDFDLAKQGVV 163

```

```

QY 344 LKD--HNDEL-----TELSNAKELKRNKDKSLSEKASKIOELE 380
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 LSDRHHQOELEKEKKVTEATKYQISEELETYKQKVESTMDLTQKQNRVSOLE 219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
Q10371
ID Q10371 PRELIMINARY; PRT; 355 AA.
AC Q10371;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE M PROTEIN, SEROTYPE 1.0 PRECURSOR (FRAGMENTS).
GN EM1.0.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS130;
RX MEDLINE=93360826; PubMed=8355619;
RA Harbaugh M.P., Podbielski A., Huegl S., Cleary P.P.;
RT "Nucleotide substitutions and small-scale insertion produce size and
RT antigenic variation in group A streptococcal M1 protein.";
RL Mol. Microbiol. 8:981-991(1993).
RN [2]
RP SEQUENCE OF 1-227 FROM N.A.
RC STRAIN=CS130;
RA Haanes-Fritz E., Kraus W., Burdett V., Dale J.B., Beachey E.H.,
RA Cleary P.;
RL Nucleic Acids Res. 16:4667-4676(1993).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
CC PHAGOCYTOSIS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
DR EMBL; X07933; CAA30757.1; -.
DR INTERPRO; IPR001899; -.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
KW Transmembrane; Coiled coil; Signal.
FT SIGNAL 1 41
FT CHAIN 42 355 M1.0 PROTEIN.
FT NON CONS 227 228 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 42 329 3 X APPROXIMATE TANDEM REPEATS, A-TYPE.
FT DOMAIN 134 227 GLY/PRO-RICH.
FT DOMAIN 284 320 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT DOMAIN 321 325 PROTEINS.
FT TRANSMEM 330 349 MEMBRANE ANCHOR.
FT DOMAIN 350 355 CYTOPLASMIC (POTENTIAL).
FT REPEAT 134 161 A-1.
FT REPEAT 162 189 A-2.
FT REPEAT 190 >227 A-3 (PARTIAL).
SQ SEQUENCE 355 AA; 39595 MW; FF839CCAECDD801B9 CRC64;

Query Match 16.4%; Score 321; DB 2; Length 355;
Best Local Similarity 46.1%; Pred. No. 2e-08;
Matches 100; Conservative 15; Mismatches 62; Indels 40; Gaps 9;

QY 200 NGDGNPREVIEDLAANNPAIONRLRHENKDKLKALENAMVAGRDFKRA-----GTL 252
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 42 NGDGNPREVIEDLAANNPAIONRLRHENKDKLKALENAMVAGRDFKRAELEKAKQAL 101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 253 LDQVTLTKHNSYQYN-AQAGRLDLRQAEYLLKGLNDWAERLIQELNIDVATRS--Q 309
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 102 EDQKDLTKLKLQDDYDLAKESTWDRQRL--KELEKKEAL--ELAIQDASRDYHR 157
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 TDITLEKQVE-----RADKFEIENNTLKLK----NSDSLSPNNKALKDHN 348
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 158 ATALEKELEKKALELAIDQASQDYNRANVLEKELETITREQENRNLGNKLELDOL 217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 349 DELTELSNAKELKRR---NDKSLSEKASKIOELEAR 382
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 SSEKEQLTIRLEKLNKELESKKLTKAK--AELOAK 252
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
Q54835
ID Q54835 PRELIMINARY; PRT; 581 AA.
AC Q54835;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE M3 PROTEIN PRECURSOR.
GN EMM3.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE M3 4 /55(PRAGUE);
RX MEDLINE=96157064; PubMed=8586172;
RA Reichardt W., Gubbe K., Schmidt K.H.;
RT "M3-protein with close sequence homology to M12 protein binds
RT fibrinogen, albumin, fibronectin, but not to any subclass of IgG-
RT localization of binding regions.";
RL Dev. Biol. Stand. 85:179-182(1995).
DR EMBL; X80168; CAA56449.1; -.
DR HSP; P03069; IGCL.
DR INTERPRO; IPR001899; -.
DR PFAM; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 41 POTENTIAL.
FT CHAIN 42 581 M3 PROTEIN.
SQ SEQUENCE 581 AA; 64958 MW; 3689317C98359D75 CRC64;

Query Match 16.4%; Score 320; DB 2; Length 581;
Best Local Similarity 44.9%; Pred. No. 3.8e-08;
Matches 79; Conservative 21; Mismatches 36; Indels 40; Gaps 6;

QY 234 RLENAMVAGRDFKRAETLLDQVTLTKHNSYQYNQAQAGRLDLRQAEYLLKGLNDWA 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 KLNKEI-----NLLDQVTLTKHNSYQYNQAQAGRLDLRQAEYLLKGLNDWA 104
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 294 ERLLOELNIDVATR-----SQDTLEK-----VQERADKFEIENNTLKLKNSDLSPNNKA 343
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 ERLLOELNGEDVKVGLGKVAPEKDDLEKEVLEKIDKEKEQDLD-KDFDLAKQGYV 163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 344 LKD--HNDEL-----TELSNAKELKRNKDKSLSEKASKIOELE 380
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 LSDRHHQOELEKEKKVTEATKYQISEELETYKQKVESTMDLTQKQNRVSOLE 219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
Q9KGR8
ID Q9KGR8 PRELIMINARY; PRT; 231 AA.
AC Q9KGR8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE M PROTEIN (FRAGMENT).
GN EMM.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]

```

```
RP SEQUENCE FROM N.A.
RC STRAIN-7203-99;
RA Beall B.W.;
RT "5" emm sequence emm3.3.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF288753; AAF91497.1;
FT NON_TER 1
FT NON_TER 231
SQ SEQUENCE 231 AA; 26305 MW; F5002430DD229351 CRC64;

Query Match 16.2%; Score 317.5; DB 2; Length 231;
Best Local Similarity 44.8%; Pred. No. 1.9e-08;
Matches 81; Conservative 20; Mismatches 35; Indels 45; Gaps 7;

Qy 234 RLENAMEVAGDRFRAGTLLDQVTL-----YTKHNSYQYNAQAGRLDLRQKAEYLGK 288
    :||:|
Db 37 LKNEIE-----NLLDQVTLTKHYTKHNSYQYNAQAGRLDLRQKAEYLGK 86

Qy 289 LNDWAERLQELNIDVATR-----SOTDTLEK----VOBRADKFEIENNTLKLKNSDLS 338
    :||:|
Db 87 LNDWAERLQELNGEDVKKVLGVAFKEDLKEVKEKIDKKEKEYQDLD-KDFDLA 145

Qy 339 FNNKALKD--HNDEL-----TEELSSNAKEKLRKNDKSLSEKASKIQEQL 379
    :||:|
Db 146 KQGVVLSDKRHOQLEKEKKVTEATAKVGQISELETVKQKVESTMODLTKQNRVSOL 205

Qy 380 E 380
Db 206 E 206

RESULT 8
Q54830 PRELIMINARY; PRT; 488 AA.
ID Q54830;
AC Q54830;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE M 3 PROTEIN PRECURSOR (FRAGMENT).
GN EM33.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=93360826; PubMed=8355619;
RA Harbaugh M.P., Podbielski A., Huegl S., Cleary P.P.;
RT "Nucleotide substitutions and small-scale insertion produce size and
antigenic variation in group A streptococcal M1 protein.";
RL Microbiol. 8:981-991(1993).
CC !- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
PHAGOCYTOSIS.
CC !- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
DR EMBL; X62131; CAA44062.1;
DR INTERPRO; IPR001899;
DR PFAM; PF00746; Gram_pos_anchor; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00143; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Virulence; Phagocytosis; Cell wall; Duplication; Repeat; Antigen;
KW Transmembrane; Coiled coil; Signal.
FT SIGNAL 1 41
FT CHAIN 42 484 M1.2 PROTEIN.
FT DOMAIN 42 458 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 134 244 3 X APPROXIMATE TANDEM REPEATS, A-TYPE.
FT DOMAIN 245 379 3 X APPROXIMATE TANDEM REPEATS, C-TYPE.
FT DOMAIN 413 449 GLY/PRO-RICH.
FT DOMAIN 450 455 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
PROTEINS.
FT TRANSMEM 459 478 MEMBRANE ANCHOR.
FT DOMAIN 479 484 CYTOPLASMIC (POTENTIAL).
FT REPEAT 134 161 A-1.
FT REPEAT 162 189 A-2.
FT REPEAT 190 244 A-3.
FT REPEAT 245 286 C-1.
FT REPEAT 287 327 C-2.
FT REPEAT 328 379 C-3.
SQ SEQUENCE 484 AA; 54234 MW; 4179A7BBAC5FAED9 CRC64;

Query Match 15.9%; Score 311; DB 2; Length 484;
Best Local Similarity 44.2%; Pred. No. 8.3e-08;
Matches 96; Conservative 22; Mismatches 61; Indels 38; Gaps 9;

Qy 200 NGDGNPREVIEDLAANNPAIONIRLHKNKDLKARLENAMVAGDRFKFA-----GTL 252
    :||:|
Db 42 NGDGNPREVIEDLAANNPAIONIRLHKNKDLKARLENAMVAGDRFKRAELEKAKAL 101

Qy 253 LDQVTLTKHYTKHNSYQYNAQAGRLDLRQKAEYLGKLNDAERLQELNIDVATRS--Q 309
    :||:|
Db 102 EDORKDLETKLEQDYDLAKESTSWDRQRL--KELEKKEAL--ELAIDQASRDYHR 157
```

FT	NON_TER	100	100	76FE8A3C33C15CF2	CRC64;	
SEQ	SEQUENCE	100 AA;	11248 MW;			
	Query Match	13.4%;	Score 261.5;	DB 2;	Length 100;	
	Best Local Similarity	72.2%;	Pred. No. 3.4e-06;			
	Matches 57;	Conservative	2;	Mismatches 13;	Indels 7;	Gaps
Qy	200	NGDGNPREVIEDLAANNPAIONIRLRHENKDKL	KARLENAM	EVAGRDFKRA	-----GTL 252	
Db	16	NGDGNPREVIEDLAANNPAIONIRLRHENKDKL	KARLENAM	EVAGRDFKRAEELKAKQAL 75		
Qy	253	LDQVTLTYTKHNSNYQQYN 271				
Db	76	EDQKDLTKLKLQDDYD 94				
RESULT 12						
Q54638		PRELIMINARY;	PRT;	100 AA.		
ID	Q54638					
AC	Q54638;	1996 (T-EMBLrel. 01, Created)				
DT	01-NOV-1996	(T-EMBLrel. 01, Last sequence update)				
DT	01-NOV-1998	(T-EMBLrel. 08, Last annotation update)				
DE	M1.6 PROTEIN (FRAGMENT).					
OS	Streptococcus pyogenes.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;					
OC	Streptococcus					
OX	NCBI_TaxID=1314;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-MGAS 2108;					
RX	MEDLINE=95172752; PubMed=7686273;					
RA	Musser J.M., Kapur V., Szeto J., Pan X., Swanson D.S., Martin D.R.;					
RT	"Genetic diversity and relationships among Streptococcus pyogenes					
RT	strains expressing serotype M1 protein: recent intercontinental spread					
RT	of a subclone causing episodes of invasive disease.";					
RL	Infect. Immun. 63:994-1003(1995).					
DR	EMBL; U20100; AAA85113.1;					
FT	NON_TER	1				
FT	NON_TER	100	100	63317108FC3A7DF2	CRC64;	
SEQ	SEQUENCE	100 AA;	11301 MW;			
	Query Match	13.3%;	Score 260.5;	DB 2;	Length 100;	
	Best Local Similarity	72.2%;	Pred. No. 3.8e-06;			
	Matches 57;	Conservative	2;	Mismatches 13;	Indels 7;	Gaps
Qy	200	NGDGNPREVIEDLAANNPAIONIRLRHENKDKL	KARLENAM	EVAGRDFKRA	-----GTL 252	
Db	16	NGDGNPREVIEDLAANNPAIONIRLRHENKDKL	KARLENAM	EVAGRDFKRAEELKAKQAL 75		
Qy	253	LDQVTLTYTKHNSNYQQYN 271				
Db	76	EDQKDLTKLKLQDDYD 94				
RESULT 13						
Q54637		PRELIMINARY;	PRT;	100 AA.		
ID	Q54637					
AC	Q54637;	1996 (T-EMBLrel. 01, Created)				
DT	01-NOV-1996	(T-EMBLrel. 01, Last sequence update)				
DT	01-NOV-1998	(T-EMBLrel. 08, Last annotation update)				
DE	M1.5 PROTEIN (FRAGMENT).					
OS	Streptococcus pyogenes.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;					
OC	Streptococcus					
OX	NCBI_TaxID=1314;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-MGAS 1666;					
RX	MEDLINE=95172752; PubMed=7686273;					

RA Musser J.M., Kapur V., Szeto J., Pan X., Swanson D.S., Martin D.R.;
 RT "Genetic diversity and relationships among Streptococcus pyogenes
 strains expressing serotype M1 protein: recent intercontinental spread
 of a subclone causing episodes of invasive disease.";
 RL Infect. Immun. 63:994-1003(1995).
 DR EMBL; U20099; AAA85112.1; -;
 FT NON_TER 1
 SQ SEQUENCE 100 AA; 11217 MW; 6332ACD6FFE7A3F2 CRC64;

Query Match 13.3%; Score 259.5; DB 2; Length 100;
 Best Local Similarity 72.2%; Pred. No. 4.2e-06;
 Matches 57; Conservative 1; Mismatches 14; Indels 7; Gaps 1;

QY 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLAKLENAMVAGRDFKRA-----GTL 252
 DB 16 NGDGNPREVIEGLAANNPAIONIRLRHENKDKLAKLENAMVAGRDFKRAEEL 75
 QY 253 LDQVTQLYTKHNSNYQYN 271
 DB 76 EDQKQDLETKELQDYD 94

RESULT 14

Q53536 PRELIMINARY; PRT; 76 AA.
 AC Q53536;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE M PROTEIN (FRAGMENT).
 GN EMM.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95377598; PubMed=7649435;
 RA Penney T.J., Martin D.R., Williams L.C., de Malmarche S.A.,
 RA Bergquist P.L.;
 RT "A single emm gene-specific oligonucleotide probe does not recognise
 all members of the Streptococcus pyogenes M type 1.";
 RL FEMS Microbiol. Lett. 130:145-149(1995).
 DR EMBL; S79254; AAB35162.1; -;
 FT NON_TER 1
 SQ SEQUENCE 76 AA; 8383 MW; EA46B64E5657A220 CRC64;

Query Match 13.2%; Score 257; DB 2; Length 76;
 Best Local Similarity 96.2%; Pred. No. 4.1e-06;
 Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLAKLENAMVAGRDFKRACTL 252
 DB 22 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLAKLENAMVAGRDFKRAEEL 74

RESULT 15

Q54543 PRELIMINARY; PRT; 79 AA.
 AC Q54543;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE M TYPE 1 (EMML) (FRAGMENT).
 GN EMM.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=M TYPE 1;
 RA Whatmore A.M.;
 RL Thesis (1993), University of Newcastle Upon Tyne, UK.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M TYPE 1;
 RX MEDLINE=95198537; PubMed=7891551;
 RA Whatmore A.M., Kapur V., Sullivan D.J., Musser J.M., Kehoe M.A.;
 RT "Non-congruent relationships between variation in emm gene sequences
 and the population genetic structure of group A streptococci.";
 RL Mol. Microbiol. 14:619-631(1994).
 DR EMBL; U11940; AAA95556.1; -;
 DR INTERPRO; IPR000886; -;
 DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 79
 SQ SEQUENCE 79 AA; 8499 MW; AE8A1FF8A4E1B711 CRC64;

Query Match 13.2%; Score 257; DB 2; Length 79;
 Best Local Similarity 96.2%; Pred. No. 4.3e-06;
 Matches 51; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 200 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLAKLENAMVAGRDFKRACTL 252
 DB 27 NGDGNPREVIEDLAANNPAIONIRLRHENKDKLAKLENAMVAGRDFKRAEEL 79

Search completed: May 10, 2001, 08:40:08
 Job time: 150 sec

us-09-151-409-16.rspt

Tue May 15 07:29:01 2001

